Statistical inference for renewal processes

F. Comte [*], C. Duval[†]

Abstract

We consider nonparametric density estimation for interarrival times density of a renewal process. If it is possible to get continuous observation of the process, then a projection estimator in an orthonormal functional basis can be built; we choose to work on $\mathbb{R}^+$ with the Laguerre basis. Nonstandard decompositions can lead to bounds on the mean integrated squared error (MISE), from which rates of convergence on Sobolev-Laguerre spaces can be deduced, when the length of the observation interval gets large. The more realistic setting of discrete time observation with sampling rate Δ is more difficult to handle. A first strategy consists in neglecting the discretization error, and under suitable conditions on Δ, an analogous MISE bound is obtained. A more precise strategy aims at taking into account the structure of the data: a deconvolution estimator is defined and studied. In that case, we work under a simplifying "dead-zone" condition. The MISE corresponding to this strategy is given for fixed Δ as well as for small Δ. In the three cases, an automatic model selection procedure is described and gives the best MISE, up to a logarithmic term. The results are illustrated through a simulation study. **September 29, 2016**

Keywords. Density deconvolution. Laguerre basis. Nonparametric estimation. Renewal processes.
AMS Classification. 62G07, 62M09

1 Introduction

1.1 Model and Observations

Let R be a renewal process. More precisely, we denote by $(T_0, T_1, \dots, T_n, \dots)$ the jump times of R, such that $(D_i := T_i - T_{i-1})_{i\geq 1}$ are i.i.d. with density τ with respect to the Lebesgue measure supported on $[0, \infty)$. The first jump time T_0 may have a different distribution τ_0. The renewal process R is a process that counts how many jumps occurred until a given time t, *i.e.*

$$R_t = \sum_{i=0}^{\infty} \mathbb{1}_{T_i \leq t}. \tag{1}$$

These processes are used to describe the occurrences of random events: for instance in seismology to modelize the occurrence of earthquakes (see *e.g.* Alvarez (2005) or Epifani *et al.* (2014)).

In this paper we are interested in estimating the density τ. We will often assume that

$$\mu := \int_0^{\infty} x\tau(x)dx < \infty. \tag{A1}$$

We consider two different sampling schemes: first, the complete observation setting, where R is continuously observed over $[0, T]$ and second, an incomplete observation setting, where R is observed at a sampling rate Δ over $[0, T]$, where Δ is either small or fixed. The continuous observation scheme, whose study reveals to be more delicate than it may first appear, will be used as a reference point for the discrete sampling scheme. Indeed, continuous time observations are more informative

[*]Sorbonne Paris Cité, Université Paris Descartes, MAP5, UMR CNRS 8145.
[†]Sorbonne Paris Cité, Université Paris Descartes, MAP5, UMR CNRS 8145.

and a procedure based on discrete observations can, at best, attains the same rates as an optimal procedure based on the continuous observations.

Estimation of the interarrival distribution for renewal processes goes back to Vardi (1982) who proposed a consistent algorithm, based on the maximization of the likelihood. It permits to estimate this distribution from the observation of K independent trajectories (see also Vardi (1989) and the generalization of Soon and Woodroofe (1996), Guédon and Cocozza-Thivent (2003) and Adekpedjou *et al.* (2010); we also refer to the review of Gill and Keiding (2010) and the references therein). Assuming that only endpoints R_t, for a given time $t > 0$, are observed and assuming a Gamma distributed interarrival distribution, Miller and Bhat (1997) proposed a parametric estimator also based on maximum likelihood techniques. However, in the aforementioned literature, the asymptotic properties of the estimators are not investigated, therefore, rates of convergence are not derived.

1.2 Continuous observation scheme

Without loss of generality we set $T_0 = 0$, or equivalently $\tau_0(dx) = \delta_0(dx)$. Suppose that R is continuously observed over $[0, T]$, namely we observe $(R_t, t \in [0, T])$. From this, we extract the observations $(D_1, \ldots, D_{R_T})$ to estimate the density τ. The counting process R_T is such that

$$T_{R_T} = \sum_{j=1}^{R_T} D_j \leq T \quad \text{and} \quad T_{R_T+1} = \sum_{j=1}^{R_T+1} D_j > T, \tag{2}$$

therefore, we are not in the classical i.i.d. density estimation problem. This implies that R_T and D_j are dependent and that the quantity D_{R_T+1} is not observed. In addition, the random number R_T of observations depends itself on the unknown density τ. Then, the statistics R_T is not ancillary. Moreover, due to this particularity, our dataset is subject to bias selection: there is a strong representation of small elapsed times D and long interarrival times are observed less often.

These issues are clearly explained in Hoffmann and Olivier (2016) who consider a related model: age dependent branching processes. Our framework can be formalized as a degenerate age dependent branching process: we study a particle with random lifetime governed by the density τ and at its death it gives rise to one other particle with a lifetime governed by the same density τ. The difference with Hoffmann and Olivier (2016), is that in their work the underlying structure of the model is a Bellmann-Harris process which has a tree representation whereas our tree contains only one branch, a case they exclude. Therefore the solutions they propose to circumvent the latter difficulties do not apply in our setting. In particular, they derive rates of convergence as functions of the Malthus parameter, which needs to be nonzero to ensure consistency. But in the Poisson process case (which is a particular renewal process) it is easy to see that this Malthus parameter is null. Therefore, in the sequel we will employ different techniques to deal with these issues.

1.3 Discrete observation scheme

Suppose now that we observe the process R over $[0, T]$ at a sampling rate Δ, namely, we observe $\big(R_{i\Delta},\ i = 1, \ldots, \lfloor T\Delta^{-1} \rfloor\big)$. This setting introduces three difficulties. Firstly, the increments $R_{i\Delta} - R_{(i-1)\Delta}$ are not independent. Secondly, they are not identically distributed. Thirdly, from the sample $\big(R_{i\Delta},\ i = 1, \ldots, \lfloor T\Delta^{-1} \rfloor\big)$ it cannot be derived a single realization of the density of interest τ.

We consider two distinct strategies. First, under some rate constraints on Δ, we show that neglecting the discretization error leads to an estimator with properties similar to the one which has access to the whole trajectory. It also permits to bypass the aforementioned difficulties.

Otherwise, if we do not wish to impose a rate condition on Δ, these difficulties need to be handled. The first difficulty is easily overcome as the dependency structure in the sample is not severe and can be treated without additional assumptions. The second problem can be circumvented by imposing a particular value for T_0 that ensures stationarity of the increments. More precisely, assuming that (A1) holds and that T_0 has density τ_0 defined by

$$\tau_0(x) = \frac{\int_x^\infty \tau(s)ds}{\mu}, \quad x \geq 0, \tag{A2}$$

the renewal process R given by (1) is stationary (see *e.g.* Lindvall (1992) or Daley and Vere-Jones (2003)). A careful study of the third difficulty leads us to conclude that we are facing a deconvolution problem where the distribution of the noise is, in general, unknown and even depends on the unknown density τ. But, we add a simplifying assumption that permits to make explicit the distribution of the noise: we assume that there exists a positive number $\Delta \geq \eta > 0$ such that $\tau(x) = 0$, $\forall x \in [0, \eta]$ (see the so-called dead-zone assumption described below). This leads to a convolution model with noise distribution corresponding to a sum of two independent uniform densities.

1.4 Main results and organization of the paper

In this paper, we propose nonparametric projection strategies for the estimation of τ, which are all based on the Laguerre basis. It is natural for $\mathbb{R}^+$-supported densities to choose a $\mathbb{R}^+$-supported orthonormal basis. Other compactly supported orthonormal basis, such as trigonometric or piecewise-polynomial basis, may also be considered provided their support can be rigorously defined. But in the discrete observation scheme, the choice of the Laguerre basis gets crucial. Indeed, deconvolution in presence of uniform noise presents specific difficulties: in the Fourier setting, it is required to divide by the characteristic function of the noise but in the present case, this Fourier transform is periodically zero. Specific solutions are needed (see Hall and Meister (2007) and Meister (2008)) which reveal to be rather difficult to implement. On the contrary, it appears that deconvolution in the Laguerre basis can be performed without restriction and is computationally easy. This tool has been proposed by Comte *et al.* (2016) and Mabon (2015) and can be applied here.

The article is organized as follows. The continuous time observation scheme is studied in Section 2, where we build a nonparametric projection estimator of τ. An upper bound on the mean integrated squared risk (MISE) is proved, from which, under additional assumptions, we can derive rates of convergence on Sobolev-Laguerre spaces, for large T. Up to logarithmic terms, these rates match the minimax rates, derived for density estimation from i.i.d. observations by Belomestny *et al.* (2016). A model selection procedure is defined and proved to lead to an automatic bias-variance compromise. The more realistic discrete time observation scheme with step Δ is considered in Section 3. Under specific conditions on Δ, the previous procedure is extended. Additional approximation terms appear in the MISE bound, which are taken into account in the model selection procedure. Removing the condition on Δ, but under an additional dead-zone assumption on the process, a Laguerre deconvolution procedure is proposed, studied and discussed. An extensive simulation Section 4 allows to illustrate all those methods for different distributions τ and when Δ is varying. Part of the results are postponed in the Appendix. A concluding Section 5 ends the paper and presents ideas for dealing with a completely general setting. Most of the proofs are deferred to Section 6.

2 Continuous time observation scheme

In this section, we assume that the process R defined by (1) is continuously observed over $[0, T]$. Thus, the jump times $(T_i)_i$ occurring in the interval are known. We recall that

$$D_i = T_i - T_{i-1}, \ i = 1, 2, \ldots \text{ with } \quad T_0 = 0$$

are subject to constraint (2). First, we describe the projection space and then, we define and study the first estimator.

2.1 The Laguerre basis

The following notations are used below. For $t, v : \mathbb{R}^+ \to \mathbb{R}$ square integrable functions, we denote the $\mathbb{L}^2$ norm and the $\mathbb{L}^2$ scalar product respectively by

$$\|t\| = \left(\int_0^\infty t(x)^2 dx\right)^{1/2} \quad \text{and} \quad \langle t, v\rangle = \int_0^\infty t(x)v(x)dx.$$

The Laguerre polynomials $(L_k)_{k\geq 0}$ and the Laguerre functions $(\varphi_k)_{k\geq 0}$ are given by

$$L_k(x) = \sum_{j=0}^{k}(-1)^j \binom{k}{j} \frac{x^j}{j!}, \quad \varphi_k(x) = \sqrt{2}L_k(2x)e^{-x}\mathbb{1}_{x\geq 0}, \quad k \geq 0.$$

The collection $(\varphi_k, k \geq 0)$ constitutes an orthonormal basis of $\mathbb{L}^2(\mathbb{R}^+)$ (that is $\langle \varphi_j, \varphi_k\rangle = \delta_{j,k}$ where $\delta_{j,k}$ is the Kronecker symbol) and is such that

$$|\varphi_k(x)| \leq \sqrt{2}, \quad \forall x \in \mathbb{R}^+, \ \forall k \geq 0.$$

For $t \in \mathbb{L}^2(\mathbb{R}^+)$ and $\forall x \in \mathbb{R}^+$, we can write that

$$t(x) = \sum_{k=0}^{\infty} a_k(t)\varphi_k(x), \quad \text{where } a_k(t) = \langle t, \varphi_k\rangle.$$

We define the m-dimensional space $S_m = \text{span}(\varphi_0, \dots, \varphi_{m-1})$ and t_m the orthonormal projection of t on S_m, we have $t_m = \sum_{k=0}^{m-1} a_k(t)\varphi_k$.

2.2 Projection estimator and upper risk bound

We are in a density estimation problem where the target density is supported on $[0, \infty)$, we assume that τ is square-integrable on $\mathbb{R}^+$ and decompose it in the Laguerre basis

$$\tau(x) = \sum_{k=0}^{\infty} a_k(\tau)\varphi_k(x), \quad x \in [0, \infty),$$

where $a_k(\tau) = \langle \varphi_k, \tau\rangle$. From this, we derive an estimator of τ based on the sample $(D_1, \dots, D_{R_T})$, defined, for $m \in \mathbb{N}$ and $x \in [0, \infty)$, by

$$\widehat{\tau}_m(x) = \sum_{k=0}^{m-1} \widehat{a}_k\varphi_k(x), \quad \text{where} \quad \widehat{a}_k = \frac{1}{R_T}\sum_{i=1}^{R_T} \varphi_k(D_i), \quad 0 \leq k \leq m-1, \tag{3}$$

where by convention $0/0 = 0$. Clearly, $\widehat{\tau}_m$ is in fact an estimator of τ_m, the orthogonal projection of τ on S_m. Since R_T is not an ancillary statistics, conditioning on the value of R_T does not simplify the study of $\widehat{a}_k$, in particular it is not possible to study easily its bias or its variance. We can bound the mean-square error of the estimator as follows.

Theorem 2.1. *Assume that $\tau \in \mathbb{L}^2(\mathbb{R}^+)$. Then, for any integer m, the estimator $\widehat{\tau}_m$ given by (3) satisfies*

$$\mathbb{E}\left[\|\widehat{\tau}_m - \tau\|^2\right] \leq \|\tau - \tau_m\|^2 + 8m\mathbb{E}\left[\frac{\mathbb{1}_{R_T\geq 1}}{R_T}\right] + \mathbf{C}_1\|\tau\|^3 \exp\left(-\frac{\kappa'}{4\sqrt{2}\|\tau\|}\sqrt{m}\right) + \mathbf{C}_2 m\sqrt{\mathbb{E}\left[\frac{\mathbb{1}_{R_T\geq 1}}{R_T^4}\right]},$$

where $\mathbf{C}_1$ and $\mathbf{C}_2$ are given in (20) and only depend on a universal constant κ'.

The bound given by Theorem 2.1 is a decomposition involving two main terms: a squared bias term, $\|\tau - \tau_m\|^2$ and a variance term $8m\mathbb{E}\left[\mathbb{1}_{R_T\geq 1}/R_T\right]$. Conditions ensuring that, for $T \leq m$, the two final terms are indeed negligible are given in the next section.

2.3 Rates of convergence

To obtain explicit rates from Theorem 2.1, we need to know the order of quantities of the form $\mathbb{E}[\mathbb{1}_{R_T\geq 1}R_T^{-\alpha}]$ for $\alpha \geq 0$. Suppose that (A1) and that the following hold: there exist positive constants σ^2 and c such that

$$\mathbb{E}[D_1^k] \leq \frac{k!}{2}\sigma^2 c^{k-2}, \qquad \forall k \geq 2. \tag{A3}$$

Assumption (A3) is a standard preliminary for applying a Bernstein inequality. It is fulfilled by Gaussian, sub-gaussian, Gamma or bounded densities. Under (A3), we can establish the following result.

Proposition 2.1. *Assume that (A1) and (A3) hold. Let $\alpha > 0$, then we have*

$$\mathbb{E}\Big[\frac{\mathbb{1}_{R_T\geq 1}}{R_T^{\alpha}}\Big] \leq \mathfrak{C}_1 T^{-\alpha}, \tag{4}$$

where $\mathfrak{C}_1 = (3\mu)^\alpha + \widetilde{C}_1$, and $\widetilde{C}_1$ is given in (24) hereafter if $\alpha \geq 1/2$ or in (25) if $\alpha \in (0, \frac{1}{2})$. If in addition T is such that $\mathbb{P}(R_T \geq 1) \geq \mathtt{a}$, $\mathtt{a} \in (0,1)$, then it also holds that:

$$\mathbb{E}\Big[\frac{\mathbb{1}_{R_T\geq 1}}{R_T^{\alpha}}\Big] \geq \mathfrak{C}_2 T^{-\alpha}, \tag{5}$$

where $\mathfrak{C}_2 = \mathtt{a}(\mu/2)^\alpha$.

Proposition 2.1 states both upper (4) and lower (5) bounds in order to control quantities of the form $\mathbb{E}[\mathbb{1}_{R_T\geq 1}R_T^{-\alpha}]$, for $\alpha > 0$. Only the upper bound is used in the sequel to compute the rates of convergence of $\widehat{\tau}_m$, but the lower bound ensures that the order in T of the upper bound is sharp.

For $s \geq 0$, the Sobolev-Laguerre space with index s (see Bongioanni and Torrea (2009), Comte and Genon-Catalot (2015)) is defined by:

$$W^s = \Big\{f : (0, +\infty) \to \mathbb{R}, f \in \mathbb{L}^2((0,+\infty)), |f|_s^2 := \sum_{k\geq 0} k^s a_k^2(f) < +\infty\Big\}.$$

where $a_k(f) = \int_0^{+\infty} f(u)\varphi_k(u)du$. For s integer, the property $|f|_s^2 < +\infty$ can be linked with regularity properties of the function f (existence of s-order derivative, but not only). We define the ball $W^s(M)$:

$$W^s(M) = \left\{f \in W^s, |f|_s^2 \leq M\right\}.$$

On this ball, we can handle the bias term $\|\tau - \tau_m\|^2$ and we obtain the following Corollary.

Corollary 2.1. *Assume that (A1) and (A3) hold and that τ belongs to $W^s(M)$. Then, for T large enough, choosing $m_{\text{opt}} = CT^{1/(s+1)}$, yields*

$$\mathbb{E}\big[\|\widehat{\tau}_{m_{\text{opt}}} - \tau\|^2\big] \leq C(M, \sigma^2, c)T^{-s/(s+1)}$$

where $C(M, \sigma^2, c)$ is a constant depending on M, σ^2, c but not on T.

Proof of Corollary 2.1. For $\tau \in W^s(M)$, we have $\|\tau - \tau_m\|^2 = \sum_{j\geq m} a_j^2(\tau) \leq Mm^{-s}$. Moreover, under (A3), we get by Inequality (4) of Proposition 2.1 that

$$8m\mathbb{E}\Big[\mathbb{1}_{R_T\geq 1}/R_T\Big] \leq Cm/T.$$

The tradeoff between these terms implies the choice $m_{\text{opt}} = CT^{1/(s+1)}$. Then we easily get that

$$\mathbf{C}_1\|\tau\|^3 \exp\Big(-\frac{\kappa'}{4\sqrt{2}\|\tau\|}\sqrt{m_{\text{opt}}}\Big) + \mathbf{C}_2 m_{\text{opt}}\sqrt{\mathbb{E}\Big[\frac{\mathbb{1}_{R_T\geq 1}}{R_T^4}\Big]} \leq \frac{\mathbf{C}_3}{T},$$

for some constant $\mathbf{C}_3$. Therefore, $\mathbb{E}\big[\|\widehat{\tau}_{m_{\text{opt}}} - \tau\|^2\big] \leq Mm_{\text{opt}}^{-s} + Cm_{\text{opt}}/T + O(1/T)$ and thus $\mathbb{E}\big[\|\widehat{\tau}_{m_{\text{opt}}} - \tau\|^2\big] \leq O(T^{s/(s+1)})$, which is the result of Corollary 2.1. □

The rate stated in Corollary 2.1 corresponds to the Sobolev-Laguerre upper bound for density estimation from T i.i.d. observations drawn in the distribution τ. This rate is proved to be minimax optimal, up to a logarithmic term, in Belomestny *et al.* (2016).

2.4 Adaptive procedure

We propose a data driven way of selecting m. For this, we proceed by mimicking the bias-variance compromise. Setting

$$\mathcal{M}_T = \{\lfloor \log^2(T) \rfloor, \lfloor \log^2(T) \rfloor + 1, \dots, \lfloor T \rfloor\},$$

where $\lfloor x \rfloor$ stands for the largest integer less than or equal to x, we select

$$\widehat{m} = \arg\min_{m \in \mathcal{M}_T} \left(-\|\widehat{\tau}_m\|^2 + \widehat{\mathrm{pen}}(m) \right) \quad \text{where} \quad \widehat{\mathrm{pen}}(m) = \kappa\big(1 + 2\log(1 + R_T)\big)\frac{m}{R_T}\mathbb{1}_{R_T \geq 1}.$$

Indeed, as $\|\tau - \tau_m\|^2 = \|\tau\|^2 - \|\tau_m\|^2$, the bias is estimated by $-\|\hat{\tau}_m\|^2$ up to the unknown but unnecessary constant $\|\tau\|^2$. On the other hand, the penalty corresponds to a random version of the variance term increased by the logarithmic term $\log(1 + R_T)$. The quantity κ is a numerical constant. In practice, κ is chosen by preliminary simulation experiments. For calibration strategies (dimension jump and slope heuristics), the reader is referred to Baudry *et al.* (2012). We prove the following result.

Theorem 2.2. *Assume that $\tau \in \mathbb{L}^2(\mathbb{R}^+)$ and $T \geq \exp(6\|\tau\|)$. Then there exists a value κ_0 such that for any $\kappa \geq \kappa_0$, we have*

$$\mathbb{E}\big[\|\widehat{\tau}_{\widehat{m}} - \tau\|^2\big] \leq c \inf_{m \in \mathcal{M}_T} \big\{\|\tau - \tau_m\|^2 + \mathbb{E}[\widehat{\mathrm{pen}}(m)]\big\} + 16c'\mathbb{E}^{1/2}\left[\frac{T^4\mathbb{1}_{R_T \geq 1}}{R_T^6}\right]$$

where c is a numerical constant ($c = 4$ would suit) and c' is defined in (33).

Compared to the result stated in Theorem 2.1, the inequality obtained in Theorem 2.2 implies that the estimator $\widehat{\tau}_{\widehat{m}}$ automatically reaches the bias-variance compromise, up to the logarithmic factor in the penalty and the multiplicative constant c. Under assumptions (A1) and (A3), the last two additional terms are negligible, if T gets large.

Rates of convergence can be derived from Theorem 2.2 by applying inequality (4) of Proposition 2.1 together with the following Corollary.

Corollary 2.2. *Assume that (A1) and (A3) hold. Then, the following holds*

$$\mathbb{E}\left[\frac{\log(1 + R_T)}{R_T}\mathbb{1}_{R_T \geq 1}\right] \leq \frac{\sqrt{\mathfrak{C}_1}}{T}\big(\mathfrak{C}_3 + \log(T + 1)\big),$$

where $\mathfrak{C}_1$ is defined in Proposition 2.1 and $\mathfrak{C}_3 = \log(2) + |\log(\mu_1)|$, with $\mu_1 = \mathbb{E}[D_1 \wedge 1]$.

Indeed, under assumptions (A1) and (A3) and if τ belongs to $W^s(M)$, the MISE $\mathbb{E}[\|\widehat{\tau}_{\widehat{m}} - \tau\|^2]$ is automatically of order $(T/\log(1 + T))^{-s/(1+s)}$, without requiring any information on τ nor s. This is the best possible rate, up to a logarithmic factor.

3 Discrete time observation scheme

In this section, we assume that only discrete time observations with step Δ, $(R_{i\Delta})_{i\Delta \in [0,T]}$ are available for estimating τ.

3.1 Observation scheme

Information about τ is brought by the position of nonzero increments. But when only discrete time observations of R over $[0, T]$ at sampling rate Δ are available, this information is partial. Indeed, let $i_0 \geq 1$ be such that $R_{i_0\Delta} - R_{(i_0-1)\Delta} \neq 0$, this entails that at least one jump occurred between $(i_0 - 1)\Delta$ and $i_0\Delta$. But,

- It is possible that more than one jump occurred between $(i_0 - 1)\Delta$ and $i_0\Delta$. However, if Δ gets small enough, the probability of this event tends to 0.

- It does not accurately determine a jump position T_i, but locates a jump time with an error bounded by 2Δ. We have no direct observations of random variables with density τ.

Consider the estimators $\widehat{T}_i^\Delta$ of the unobserved jump times defined recursively by

$$\widehat{T}_0^\Delta = \min\{k > 0,\ R_{k\Delta} - R_{(k-1)\Delta} \neq 0\} \times \Delta$$
$$\widehat{T}_i^\Delta = \min\{k > \tfrac{1}{\Delta}\widehat{T}_{i-1}^\Delta,\ R_{k\Delta} - R_{(k-1)\Delta} \neq 0\} \times \Delta, \qquad i \geq 1.$$

To estimate τ, we use the observations

$$(\widehat{D}_i^\Delta := \widehat{T}_i^\Delta - \widehat{T}_{i-1}^\Delta,\ i = 1, \ldots, N_T)$$

where $N_T = \sum_{i=1}^{\lfloor T\Delta^{-1}\rfloor} \mathbb{1}_{R_{i\Delta} \neq R_{(i-1)\Delta}}$ is the random number of observed nonzero increments. We drop the observation $\widehat{T}_0^\Delta$ since it is related to the density τ_0 and not τ.

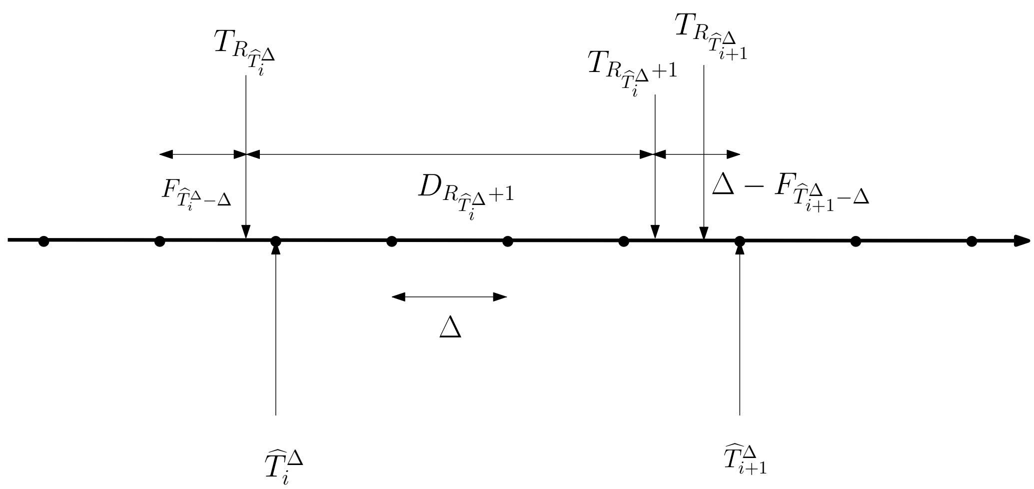

Figure 1: Discrete time observation scheme.

Let $i \geq 0$, Figure 1 illustrates how the observation $\widehat{D}_{i+1}^\Delta = \widehat{T}_{i+1}^\Delta - \widehat{T}_i^\Delta$ is related to the other quantities at stake. Define $F_t = \min\{T_j - t, \forall j, T_j \geq t\}$, the forward time at time t: that is the elapsed time from t until the next jump. By definition of R and the forward times, the following equality holds: $\widehat{D}_{i+1}^\Delta + \Delta = D_{R_{\widehat{T}_i^\Delta}+1} + F_{\widehat{T}_i^\Delta-\Delta} + (\Delta - F_{\widehat{T}_{i+1}^\Delta-\Delta})$, leading to

$$\widehat{D}_{i+1}^\Delta = D_{R_{\widehat{T}_i^\Delta}+1} + F_{\widehat{T}_i^\Delta-\Delta} - F_{\widehat{T}_{i+1}^\Delta-\Delta}. \tag{6}$$

Equation (6) shows that the observable quantity $\widehat{D}_{i+1}^\Delta$ is the sum of one realization of τ, $D_{R_{\widehat{T}_i^\Delta}+1}$, plus an error term given by $F_{\widehat{T}_i^\Delta-\Delta} - F_{\widehat{T}_{i+1}^\Delta-\Delta}$. Moreover, using the renewal property, which ensures that trajectories separated by jump times are independent, we derive that $D_{R_{\widehat{T}_i^\Delta}+1}$, $F_{\widehat{T}_i^\Delta-\Delta}$ and $F_{\widehat{T}_{i+1}^\Delta-\Delta}$ are independent. Therefore, we recover a deconvolution framework. However, for consecutive indices, the observations $\widehat{D}_i^\Delta$ and $\widehat{D}_{i+1}^\Delta$ are dependent since they both depend on the variable $F_{\widehat{T}_i^\Delta-\Delta}$. An issue that is easily circumvented by considering separately odd and even indices.

In the following, we consider observations $\widehat{D}_i^\Delta$ as given in (6) and we denote by f_Δ the density of the $\widehat{D}_i^\Delta$'s. In Section 3.2, we prove that $\widehat{D}_i^\Delta = D_i' + F_{\widehat{T}_i^\Delta-\Delta} - F_{\widehat{T}_{i+1}^\Delta-\Delta}$, with (D_i') i.i.d. with density τ and study the impact of neglecting the term $F_{\widehat{T}_i^\Delta-\Delta} - F_{\widehat{T}_{i+1}^\Delta-\Delta}$. In Section 3.3, we take the complete structure into account but we add a "dead-zone" assumption (A4) given below, that allows to compute the density of $F_{\widehat{T}_i^\Delta-\Delta} - F_{\widehat{T}_{i+1}^\Delta-\Delta}$. We can then consider a deconvolution strategy.

3.2 A first naive but general procedure

In this Section, we investigate a procedure which neglects the observation bias. For small Δ, this corresponds to the approximation $f_\Delta \asymp \tau$. Using again the decomposition of the density τ in the Laguerre basis, we define an estimator of τ based on the sample $(\widehat{D}_1^\Delta, \ldots, \widehat{D}_{N_T}^\Delta)$, by setting, for $m \in \mathbb{N}$ and $x \in [0, \infty)$

$$\check{\tau}_m(x) = \sum_{k=0}^{m-1} \check{a}_k \varphi_k(x), \quad \text{where} \quad \check{a}_k = \frac{1}{N_T} \sum_{i=1}^{N_T} \varphi_k(\widehat{D}_i^\Delta), \quad 0 \leq k \leq m-1. \tag{7}$$

Starting from (6), we can prove the following Lemma.

Lemma 3.1. *We have*

$$\widehat{D}_i^\Delta = D_{R_{\widehat{T}_i^\Delta}+1} + \Delta \xi_i, \quad 1 \leq i \leq N_T, \tag{8}$$

where $D_{R_{\widehat{T}_i^\Delta}+1}$ are i.i.d. with density τ and (ξ_i) are random variables taking values in $[-1, 1]$.

Thanks to Lemma 3.1, we can bound the mean-squared error of the estimator as follows.

Proposition 3.1. *Assume that $\tau \in \mathbb{L}^2(\mathbb{R}^+)$. Then, for any integer m, the estimator $\check{\tau}_m$ given by (7) satisfies*

$$\mathbb{E}\big[\|\check{\tau}_m - \tau\|^2\big] \leq \|\tau - \tau_m\|^2 + 16m\mathbb{E}\Big[\frac{\mathbb{1}_{N_T \geq 1}}{N_T}\Big] + 2\mathbf{C}_1 \|\tau\|^3 \exp\Big(-\frac{\kappa'}{4\sqrt{2}\|\tau\|}\sqrt{m}\Big)$$
$$+ 2\mathbf{C}_2 m \sqrt{\mathbb{E}\Big[\frac{\mathbb{1}_{N_T \geq 1}}{N_T^4}\Big]} + \Delta^2 \frac{m(4m^2-1)}{3},$$

where $\mathbf{C}_1$ and $\mathbf{C}_2$ are given in (20) and only depend on a universal constant κ'.

The result of Proposition 3.1 completes the bound obtained in Theorem 2.1: R_T is replaced by N_T and an additional error term of order $\Delta^2 m^3$, due to the model approximation appears in the bound. It is small only if Δ is small. Using the result stated in inequality (4) of Proposition 2.1, we obtain the following Corollary, which gives a condition under which the rate corresponding to the continuous time observation scheme is preserved.

Corollary 3.1. *Assume that (A1) and (A3) hold, that τ belongs to $W^s(M)$ and that $R_T = N_T$ a.s. Then for T large enough and Δ such that $\Delta^2 T^3 \leq 1$, choosing $m_{\text{opt}} = CT^{1/(s+1)}$, yields*

$$\mathbb{E}\big[\|\check{\tau}_{m_{\text{opt}}} - \tau\|^2\big] \leq C(M, \sigma^2, c) T^{-s/(s+1)}$$

where $C(M, \sigma^2, c)$ is a constant depending on M, σ^2, c but not on T.

Indeed, the additional term compared to Corollary 2.1 is $\Delta^2 m(4m^2-1)/3 \leq C\Delta^2 m^3 \leq \Delta^2 m T^2$, as $m \leq T$. Therefore, we have $\Delta^2 m T^2 \leq m/T$ if $\Delta^2 T^3 \leq 1$.

Remark 3.1. Note that $R_T = N_T$ a.s. is satisfied under Assumption (A4) below. In addition, we emphasize that we can obtain Corollary 3.1 by replacing the assumption $R_T = N_T$ a.s. by the assumption $\forall x \geq 0, \tau(x) \leq \beta_1 \exp(-\beta_2 x^{\beta_3})$ where $\beta_1, \beta_2, \beta_3$ are positive constants. Indeed, under this condition, the result of Lemma 7.3 in Duval (2013b) allows to obtain inequality (4) of Proposition 2.1 with R_T replaced by N_T.

For model selection, the procedure studied in Theorem 2.2 can be extended as follows. We define

$$\check{m} = \arg\min_{m \in \mathcal{M}_T} \big(-\|\check{\tau}_m\|^2 + \check{\mathrm{pen}}(m)\big) \quad \check{\mathrm{pen}}(m) = \Big(\check{\kappa}_1\big(1 + 2\log(1+N_T)\big)\frac{m}{N_T}\mathbb{1}_{N_T \geq 1} + \check{\kappa}_2 \Delta^2 m^3\Big),$$

where $\mathcal{M}_T$ is as previously. Then we can prove the following result

Theorem 3.1. *Assume that $\tau \in \mathbb{L}^2(\mathbb{R}^+)$ and $T \geq e^{6\|\tau\|}$. Then there exists a value $\breve{\kappa}_0$ such that for any $\breve{\kappa}_1$, $\breve{\kappa}_2$, $\breve{\kappa}_1 \vee \breve{\kappa}_2 \geq \breve{\kappa}_0$, we have*

$$\mathbb{E}\left[\|\breve{\tau}_{\breve{m}} - \tau\|^2\right] \leq \breve{c} \inf_{m \in \mathcal{M}_T} \left\{\|\tau - \tau_m\|^2 + \mathbb{E}[\breve{\text{pen}}(m)]\right\} + \breve{c}'' \mathbb{E}^{1/2}\left[\frac{T^4 \mathbb{1}_{N_T \geq 1}}{N_T^6}\right]$$

where $\breve{c}$ and $\breve{c}''$ are numerical constants ($\breve{c} = 4$ would suit).

If $\Delta^2 T^3 \leq 1$, the remarks made after Theorem 2.2 still apply here (see also the numerical Section 4).

3.3 Case of a dead-zone

3.3.1 The dead-zone assumption

Our dead-zone assumption is the following:

$$\exists \eta > 0, \quad \tau(x) = 0, \quad \forall x \in [0, \eta] \text{ with } \Delta < \eta. \tag{A4}$$

In other words when a jump occurs, no jump can occur in the next η units of times. Then, for $\Delta < \eta$, we have $\mathbb{P}(R_\Delta > 1 | R_\Delta \neq 0) = 0$ and clearly $N_T = R_T$ a.s. Moreover, the decomposition (6) becomes then

$$\widehat{D}_{i+1}^\Delta = D_{i+1} + F_{\widehat{T}_i^\Delta - \Delta} - F_{\widehat{T}_{i+1}^\Delta - \Delta}, \quad i \geq 1, \tag{9}$$

and we denote by g_Δ the density of $F_{\widehat{T}_i^\Delta - \Delta}$. The following key property holds.

Lemma 3.2. *Assume that (A1), (A2) and (A4) hold. Then, D_i, $F_{\widehat{T}_i^\Delta - \Delta}$ and $F_{\widehat{T}_{i+1}^\Delta - \Delta}$ are independent and $F_{\widehat{T}_i^\Delta - \Delta}$ and $F_{\widehat{T}_{i+1}^\Delta - \Delta}$ have common density g_Δ, equal to the uniform distribution on $[0, \Delta]$.*

Therefore, the density f_Δ of the observations $(\widehat{D}_i^\Delta)_{i \geq 1}$ as given in (9) can be written

$$f_\Delta := \tau * g_\Delta * g_\Delta(-.)(x) \quad \text{where} \quad g_\Delta * g_\Delta(-.)(x) = \frac{\Delta - |x|}{\Delta^2} \mathbb{1}_{[-\Delta, \Delta]}(x), \quad x \in \mathbb{R}. \tag{10}$$

Since we use Laguerre basis decomposition, we need the distribution of the error $g_\Delta * g_\Delta(-.)$ to be supported on $(0, \infty)$. This is why we transform the observations as follows

$$Y_i^\Delta := \widehat{D}_i^\Delta + \Delta \stackrel{d}{=} D_i + \Delta(U_i + V_i), \quad 1 \leq i \leq R_T, \tag{11}$$

where $\stackrel{d}{=}$ means equality in law and (U_i) and (V_i) are independent and i.i.d. with distribution $\mathcal{U}[0, 1]$. The density of Y_i^Δ follows from (10) and is $f_\Delta(. - \Delta)$.

3.3.2 Preliminary remark about Fourier deconvolution

Let us briefly discuss why it is not relevant to use here the classical Fourier strategy. Let $\mathcal{F}[h](u) = \int_\mathbb{R} e^{iux} h(x) dx$ denote the Fourier transform of an integrable function h. Then, under assumption (A4), we get, for all $u \in \mathbb{R}$

$$\mathcal{F}[f_\Delta](u) = \int_\mathbb{R} e^{iux} (\tau * g_\Delta * g_\Delta(-.))(x) dx = \mathcal{F}[\tau](u) \left|\mathcal{F}[g_\Delta](u)\right|^2 = \mathcal{F}[\tau](u) \times \frac{\left(\sin(\frac{u\Delta}{2})\right)^2}{\left(\frac{u\Delta}{2}\right)^2}.$$

We can see that recovering $\mathcal{F}[\tau](u)$ (and then τ by Fourier inversion) would require to divide by a sinusoidal function which can be zero. The general Fourier deconvolution setting excludes such possibility (see *e.g.* Fan (1991)). However, the case of oscillating Fourier transforms of the noise has been studied (see Hall and Meister (2007) and Meister (2008)): it is worth stressing that it requires specific methods which do not seem easy to implement. Moreover, in these papers, if the use of cross-validation techniques are suggested to achieve adaptivity, from a theoretical viewpoint this question remains open. This is why the Laguerre basis appears as an adequate answer to our problem.

3.3.3 Laguerre deconvolution

We are in a density estimation problem where the target density is supported on $[\eta, \infty)$, $\eta > 0$. However, the observations (Y_j^Δ), with density $f_\Delta(. - \Delta)$ are blurred realizations of τ, there is an additive noise supported on $[0, 2\Delta]$. We decompose the density $f_\Delta(. - \Delta)$ in the Laguerre basis

$$f_\Delta(x - \Delta) = \sum_{k=0}^{\infty} b_k \varphi_k(x), \quad x \in [0, \infty),$$

where $b_k = \langle \varphi_k, f_\Delta(. - \Delta)\rangle$. Thus, we have estimators for the b_k's, for $m \in \mathbb{N}$, defined as previously by

$$\widetilde{b}_k = \frac{1}{R_T} \sum_{i=1}^{R_T} \varphi_k(Y_i^\Delta), \quad 0 \le k \le m-1.$$

However, we are not interested in estimating $f_\Delta(. - \Delta)$ but τ. Using (11), we have that $f_\Delta = \tau * g_{2,\Delta}$ where $g_{2,\Delta}$ denotes the density of $\Delta(U_1 + V_1)$. Note that $g_{2,\Delta} = g_\Delta * g_\Delta$ where g_Δ denotes the density of ΔU_1.

The Laguerre basis has already been used in deconvolution setting by Comte *et al.* (2016) and Mabon (2015) and allows to solve the estimation problem as follows. Denoting by b_k, a_k and $g_{2,k}(\Delta)$ the coefficients of $f_\Delta(. - \Delta)$, τ and $g_{2,\Delta}$ in the Laguerre basis and plugging these expansions into the convolution, we obtain the following equation

$$\sum_{k=0}^{\infty} b_k \varphi_k(t) = \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} a_k g_{2,j}(\Delta) \int_0^t \varphi_k(x) \varphi_j(t-x) dx. \tag{12}$$

The relation (see, *e.g.* 7.411.4 in Gradshtein and Ryzhik (1980))

$$\int_0^t \varphi_k(x)\varphi_j(t-x)dx = 2e^{-t} \int_0^t L_k(2x) L_j(2(t-x))dx = 2^{-1/2}\left[\varphi_{k+j}(t) - \varphi_{k+j+1}(t)\right],$$

implies that equation (12) can be re-written

$$\sum_{k=0}^{\infty} b_k \varphi_k(t) = \sum_{k=0}^{\infty} [\sum_{\ell=0}^{k} 2^{-1/2}\left(g_{2,k-\ell}(\Delta) - g_{2,k-\ell-1}(\Delta)\right) a_\ell] \varphi_k(t),$$

with convention $g_{2,-1}(\Delta) = 0$. Equating coefficients for each of the functions, we obtain an infinite triangular system of linear equations. The triangular structure allows to increase progressively the dimension of the developments without changing the beginning.

Finally, we relate the theoretical vector $\mathbf{a}_m = (a_k)_{0 \le k \le m-1}$ of the first coefficients of decomposition of τ in the Laguerre basis with the vector $\mathbf{b}_m = (b_k)_{0 \le k \le m-1}$ as follows

$$\mathbf{b}_m = [\mathbf{G}_m(\Delta)]^2 \mathbf{a}_m,$$

where $\mathbf{G}_m(\Delta)$ is known and is the lower triangular Toeplitz matrix with elements

$$[\mathbf{G}_m(\Delta)]_{i,j} = \begin{cases} \sqrt{2}^{-1} g_0(\Delta) & \text{if } i = j \\ \sqrt{2}^{-1}\big(g_{i-j}(\Delta) - g_{i-j-1}(\Delta)\big) & \text{if } j < i \\ 0 & \text{otherwise} \end{cases} \quad \text{where } g_k(\Delta) = \int_0^\Delta \frac{1}{\Delta} \varphi_k(u) du.$$

Note that

$$[\mathbf{G}_{2,m}(\Delta)]_{i,j} = \begin{cases} \sqrt{2}^{-1} g_{2,0}(\Delta) & \text{if } i = j \\ \sqrt{2}^{-1}\big(g_{2,i-j}(\Delta) - g_{2,i-j-1}(\Delta)\big) & \text{if } j < i \\ 0 & \text{otherwise} \end{cases} \quad \text{where } g_{2,k}(\Delta) = \langle g_{2,\Delta}, \varphi_k \rangle$$

satisfies $\mathbf{G}_{2,m}(\Delta) = [\mathbf{G}_m(\Delta)]^2$. Also, we emphasize that

$$\det(\mathbf{G}_m(\Delta)) = 2^{-m/2} g_0(\Delta)^m = [(1-e^{-\Delta})/\Delta]^m > 0$$

for all Δ, which means that the matrix can be inverted. Then, we propose the following estimator of $\mathbf{a}_m$

$$\widetilde{\mathbf{a}}_m := [\mathbf{G}_m(\Delta)^2]^{-1}\widetilde{\mathbf{b}}_m.$$

This leads to the estimator of τ, for $x \geq 0$,

$$\widetilde{\tau}_m(x) = \sum_{k=0}^{m-1} \widetilde{a}_k \varphi_k(x). \tag{13}$$

3.3.4 Upper risk bound and adaptive procedure

Denote by $\rho(\mathbf{A})$ the spectral norm of a matrix $\mathbf{A}$ defined as $\rho(\mathbf{A}) = \sqrt{\lambda_{\max}(\mathbf{A}^T\mathbf{A})}$, the square-root of the largest eigenvalue of the semi definite positive matrix $\mathbf{A}^T\mathbf{A}$.

Proposition 3.2. *Assume that (A1), (A2) and (A4) hold and that $\tau \in \mathbb{L}^2(\mathbb{R}^+)$. Then, for any integer m and $\Delta \leq \eta$, the estimator $\widetilde{\tau}_m$ given by (13) satisfies*

$$\mathbb{E}\big[\|\widetilde{\tau}_m - \tau\|^2\big] \leq \|\tau - \tau_m\|^2 + \rho^2\big(\mathbf{G}_m(\Delta)^{-2}\big) 8m\mathbb{E}\Big[\frac{\mathbb{1}_{R_T\geq 1}}{R_T}\Big]$$
$$+ \rho^2\big(\mathbf{G}_m(\Delta)^{-2}\big)\left(\mathbf{C}_1\|\tau\|^3 \exp\Big(-\frac{\kappa'}{4\sqrt{2}\|\tau\|}\sqrt{m}\Big) + \mathbf{C}_2 m\sqrt{\mathbb{E}\Big[\frac{\mathbb{1}_{R_T\geq 1}}{R_T^4}\Big]}\right),$$

where $\mathbf{C}_1$ and $\mathbf{C}_2$ are given in (20) and only depend on a universal constant κ'.

Proposition 3.2 shows that the bias term is unchanged, but all other terms are multiplied by $\rho^2\big(\mathbf{G}_m(\Delta)^{-2}\big)$, which is a classical price for solving the inverse problem. In accordance with this, consider the collection

$$\widetilde{\mathcal{M}}_T = \{m \in \{\lfloor \log^2(T)\rfloor, \lfloor \log^2(T)\rfloor + 1, \dots, [T]\}, \quad m\rho^2\big(\mathbf{G}_m(\Delta)^{-2}\big) \leq T\}$$

and the selection device

$$\widetilde{m} = \arg\min_{m\in\widetilde{\mathcal{M}}_T} (-\|\widetilde{\tau}_m\|^2 + \widetilde{\mathrm{pen}}(m)), \quad \widetilde{\mathrm{pen}}(m) = \log(1+R_T)\frac{m}{R_T}\big(\widetilde{\kappa}_1 + \widetilde{\kappa}_2\rho^2\big(\mathbf{G}_m(\Delta)^{-2}\big)\big)\mathbb{1}_{R_T\geq 1}.$$

We can prove

Theorem 3.2. *Assume that (A1), (A2) and (A4) hold and $\tau \in \mathbb{L}^2(\mathbb{R}^+)$. Let $T \geq e^{8\|\tau\|}$ and $\Delta \leq \eta$. Then, there exists a value $\widetilde{\kappa}_0$, such that for any $\widetilde{\kappa}_1, \widetilde{\kappa}_2$, $\widetilde{\kappa}_1 \vee \widetilde{\kappa}_2 \geq \widetilde{\kappa}_0$, we have*

$$\mathbb{E}[\|\widetilde{\tau}_{\widetilde{m}} - \tau\|^2] \leq c_1 \inf_{m\in\widetilde{\mathcal{M}}_T} \big\{\|\tau - \tau_m\|^2 + \mathbb{E}[\widetilde{\mathrm{pen}}(m)]\big\} + c_2\mathbb{E}^{1/2}\left[\frac{T^6\mathbb{1}_{R_T\geq 1}}{R_T^8}\right] \tag{14}$$

where c_1 and c_2 are numerical constants ($c_1 = 4$ would suit).

The result of Theorem 3.2 shows that the procedure leads to the adequate squared-bias variance compromise. Under Assumption (A3), we get by Inequality (4) of Proposition 2.1 that the last term in (14) is of order $1/T$ and thus is negligible.

3.3.5 Some remarks

First, the following lemma shows that the matrix $\mathbf{G}_m(\Delta)$ is easy to compute recursively from the Laguerre basis. Therefore, formula (15) and (16), and consequently our estimator $\widetilde{\tau}_m$, can be easily implemented.

Lemma 3.3. *We have, for $k \in \mathbb{N}$,*

$$g_k(\Delta) = \frac{1}{\Delta}\big((-1)^k\sqrt{2} - \Phi_k(\Delta)\big), \text{ with } \Phi_k(\Delta) = \int_\Delta^\infty \varphi_k(u)du. \tag{15}$$

Moreover, $\forall x \in \mathbb{R}^+$, we have $\Phi_0(x) = \varphi_0(x)$ (initialization) and for $j \geq 1$, j integer,

$$\Phi_j(x) = \varphi_j(x) - \varphi_{j-1}(x) - \Phi_{j-1}(x). \tag{16}$$

Second, to compute the rate of convergence implied by Theorem 3.2, the knowledge of the spectral norm $\rho^2\big(\mathbf{G}_m(\Delta)^{-2}\big)$ is required. When Δ tends to 0 it is straightforward to observe that for all k, $\lim_{\Delta\to 0} g_k(\Delta) = \varphi_k(0) = \sqrt{2}$. It follows that $\mathbf{G}_m(\Delta) \to Id_m$, when $\Delta \to 0$, where Id_m is the $m \times m$ identity matrix. More precisely, we can get the following development

$$\mathbf{G}_m(\Delta)^{-2}[\mathbf{G}_m(\Delta)^{-2}]^T = Id_m + 2\Delta A + o(\Delta)$$

where A is the $m \times m$ matrix with all its coefficients equal to 1. This implies that $\rho^2\big(\mathbf{G}_m(\Delta)^{-2}\big)$ tends to 1 when Δ tends to 0.

For fixed Δ we propose a conjecture motivated by numerical experiments. We observe numerically that $\rho^2\big(\mathbf{G}_m(\Delta)^{-2}\big) \asymp m^4$. If this is true, the rate of the estimator is $O(T^{-s/(s+5)})$, with a logarithmic loss for the adaptive procedure. It is not clear if this rate is optimal. Indeed, in the case of T i.i.d. observations of variables blurred with additive noise of known density, the result in Mabon (2015) would give a variance term in the upper bound of order

$$\frac{1}{T}\left\{[m\rho^2\big(\mathbf{G}_m(\Delta)^{-2}\big)] \wedge [\|f_\Delta\|_\infty \|\mathbf{G}_m(\Delta)^{-2}\|_F^2]\right\}$$

where $\|\mathbf{A}\|_F^2 = \mathrm{Tr}(\mathbf{A}\mathbf{A}^T)$ denotes the Frobenius norm of the matrix $\mathbf{A}$. In the cases where the orders of the operator norm and the Frobenius norm are obtained, they turn out to be the same (see Comte *et al.* (2016)). It implies that the variance order may be governed by $\|\mathbf{G}_m(\Delta)^{-2}\|_F^2/T$ and may lead, in the case where Δ is fixed, to a better rate than the one obtained in Theorem 3.2. Nevertheless, the differences between the two terms, if any, vanishes when Δ gets small, as $m\rho^2(I_m) = \|I_m\|_F^2 = m$. However, obtaining an upper bound for the variance term involving $\|\mathbf{G}_m(\Delta)^{-2}\|_F^2/T$ is much more involved in this case than in the context considered in Mabon (2015) due to the fact that our number of observations is random and is not ancillary. Also it is difficult to compare the bound derived from Theorem 3.2, with the optimal rate derived in Meister (2008) since the regularity assumptions on the target density τ are different.

4 Simulations

In this section, we illustrate the performances of the estimators, with data driven selection of the dimension, on simulated data. We consider the following different $\mathbb{R}^+$-supported densities τ

- a Gamma $\mathcal{G}(2, \frac{1}{2})$,
- the absolute value of a Gaussian $|\mathcal{N}(1, \frac{1}{2})|$,
- a dilated Beta $5 \times \mathcal{B}(6, 3)$,
- or a rescaled mixture of Gamma densities $\big(0.4\mathcal{G}(2, \frac{1}{2}) + 0.6\mathcal{G}(16, \frac{1}{4})\big) \times \frac{8}{5}$.

The last two densities are rescaled so that for all the examples the mass is mainly contained in the interval $[0, 5]$. To estimate the $\mathbb{L}^2$-risks, we compute 1000 trajectories for $T = 500, 1000$ and 5000. The dimension m is selected among all dimensions smaller than 50. All methods require the calibration of constants in penalties. After preliminary experiments, κ is taken equal to 0.13 for the estimator based on continuous observations ($\Delta = 0$), $\breve{\kappa}_1 = 0.13$ and $\breve{\kappa}_2 = 0.001$ for the naive estimator, $\widetilde{\kappa}_1 = 0.25$ and $\widetilde{\kappa}_2 = 0.0001$ for the dead-zone estimator, which are based on discrete observations, whatever the value of nonzero Δ.

In the sequel, the different estimators are always computed on the same trajectory, even when the value of Δ is varying. Moreover, together with the value of the $\mathbb{L}^2$-risk, we provide the quantity $\overline{m}$, which is the average of dimensions $\widehat{m}$ that have been adaptively selected by each procedure and the quantity $\overline{R}$ which is the average number of observations that have been used to estimate τ. Standard deviations associated with these means are given in parenthesis. Only one distribution is presented in this Section, the other tables for the other distributions can be found in the Appendix. We present illustrations of the methods in Figures 2 and 3, which plot beams of 50 estimators computed with the three adaptive procedures, the one based on continuous observations of (R_t) as in Section 2 for $T = 500$, the ones based on discrete observations using the naive or the deconvolution method, for two different steps of observations ($\Delta = 0.3$ and $\Delta = 0.1$ in Figure 2 and $\Delta = 0.1$ and $\Delta = 0.01$ in Figure 3). We work here under the dead-zone assumption ($\eta = 1$) to permit the comparison. As expected, the procedure based on continuous time observations is very good, and the best one, but the two other methods perform also very well, even if the naive method requires smaller steps of observation.

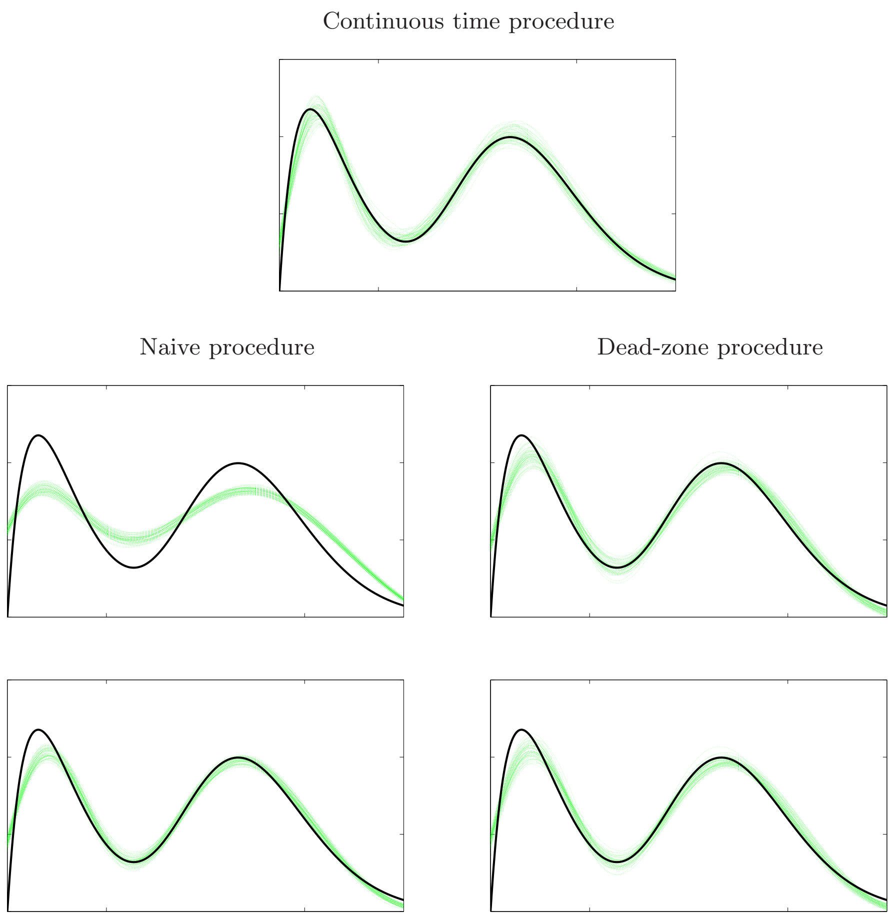

Figure 2: Estimation of τ, a shifted ($\eta = 1$) mixture of Gamma densities $\big(0.4\mathcal{G}(2, \frac{1}{2}) + 0.6\mathcal{G}(16, \frac{1}{4})\big) \times \frac{8}{5}$, for $T = 500$. The estimator based on the continuous observation (first line), $\Delta = 0.3$ (second line) and for $\Delta = 0.1$ (third line), with the naive method (first column) and the dead-zone method (second column). True density τ in black and 50 estimated curves in green.

Comparison of the continuous time and the naive procedure. The results of Table 1 confirm the theoretical results established in the paper. As expected, we notice that the best estimator is the one which has access to the continuous time observations ($\Delta = 0$). When Δ gets too large, the naive procedure is biased and performs badly. However, its performances are better in practice than what the theory predicts: even when $m^3\Delta^2$ is larger than one, the performances of the naive method are satisfactory. But when $m^3\Delta^2$ becomes too large, the method fails. Finally we recover that the larger T, the smaller the loss. The performances of the procedures are only marginally influenced by the choice for the distribution τ (see Tables 3, 4 and 5 for the other distributions in the Appendix).

T		$\Delta = 0$	$\Delta = 0.01$	$\Delta = 0.1$	$\Delta = 0.3$
500	$\overline{m}^3\Delta^2$		0.01	1.31	12.80
	$\mathbb{L}_2$	$2.42 \cdot 10^{-3}$ $(1.90 \cdot 10^{-3})$	$2.44 \cdot 10^{-3}$ $(1.90 \cdot 10^{-3})$	$2.43 \cdot 10^{-3}$ $(1.95 \cdot 10^{-3})$	$7.82 \cdot 10^{-3}$ $(2.83 \cdot 10^{-3})$
	$\overline{R}$	498.55 (15.92)	497.52 (15.91)	494.54 (15.83)	474.85 (14.65)
	$\overline{m}$	4.98 (0.67)	4.96 (0.67)	5.07 (0.70)	5.22 (0.55)
1000	$\overline{m}^3\Delta^2$		0.01	1.54	13.79
	$\mathbb{L}_2$	$1.22 \cdot 10^{-3}$ $(0.91 \cdot 10^{-3})$	$1.22 \cdot 10^{-3}$ $(0.92 \cdot 10^{-3})$	$1.33 \cdot 10^{-3}$ $(1.33 \cdot 10^{-3})$	$6.97 \cdot 10^{-3}$ $(1.74 \cdot 10^{-3})$
	$\overline{R}$	998.00 (22.13)	996.92 (22.14)	990.93 (22.01)	992.30 (20.33)
	$\overline{m}$	5.31 (0.60)	5.31 (0.60)	5.36 (0.61)	5.35 (0.49)
5000	$\overline{m}^3\Delta^2$		0.02	2.20	17.55
	$\mathbb{L}_2$	$0.27 \cdot 10^{-3}$ $(0.20 \cdot 10^{-3})$	$0.27 \cdot 10^{-3}$ $(0.20 \cdot 10^{-3})$	$0.42 \cdot 10^{-3}$ $(0.26 \cdot 10^{-3})$	$6.00 \cdot 10^{-3}$ $(0.92 \cdot 10^{-3})$
	$\overline{R}$	4998.0 (50.60)	4996.6 (50.60)	4967.0 (50.20)	4772.4 (46.30)
	$\overline{m}$	6.08 (0.43)	6.07 (0.43)	6.04 (0.33)	5.80 (0.40)

Table 1: Simulation results for τ following a $\mathcal{G}(2, \frac{1}{2})$ distribution. $\mathbb{L}_2$: mean square errors, $\overline{R}$: mean of the number of observations, $\overline{m}$: mean of the selected dimensions. All standard deviations are given in parenthesis.

Comparison of the continuous time and the dead-zone procedure. To apply the dead-zone procedure, we shifted all four distributions of a factor $\eta = 1$. We computed $\mathbb{L}^2([\eta, \infty))$ losses and compared the first and third estimators. Again, the results of Table 2 illustrate the theoretical properties established in the paper. The larger Δ, the more difficult the estimation problem is: the risks increase with Δ. But this procedure permits to consistently estimate τ even when Δ does not go to 0, whereas the latest naive procedure failed to estimate τ in theses cases. The performance of the procedure is only marginally influenced by the choice for the distribution τ (see Tables 6, 7 and 8 for the other distributions in the Appendix). Note that, for the same values of T, since the distributions have been shifted with a parameter 1, the effective number of observations $\overline{R}$ available for the estimation is smaller.

5 Concluding remarks

In this paper we propose procedures to estimate the interarrival density of a renewal process. In the case where the process is continuously observed, our procedure is adaptive minimax and requires few assumptions on the target density. The main difficulty of the problem was to deal with the random number of observation that is non ancillary. If the process is discretely observed, the problem becomes much more involved, the observations are not independent nor identically distributed and the estimation problem is of deconvolution type. When Δ goes rapidly to zero, we show that the estimation problem can be handled similarly to the estimation problem from continuous observation with preserved properties. Otherwise, we imposed additional simplifying assumptions (A1), (A2) to ensure stationarity of the increments and (A4) to manage the distribution of the noise. An adaptive procedure is proposed even though its optimality remains an open question. The numerical study confirms these theoretical considerations.

T		$\Delta = 0$	$\Delta = 0.01$	$\Delta = 0.1$	$\Delta = 0.5$	$\Delta = 0.75$
500	$\mathbb{L}_2$	$11.24 \cdot 10^{-3}$	$16.67 \cdot 10^{-3}$	$16.77 \cdot 10^{-3}$	$24.55 \cdot 10^{-3}$	$38.56 \cdot 10^{-3}$
		$(1.91 \cdot 10^{-3})$	$(1.90 \cdot 10^{-3})$	$(1.87 \cdot 10^{-3})$	$(1.57 \cdot 10^{-3})$	$(\cdot 10^{-3})$
	$\overline{R}$	248.34 (5.63)	247.34 (5.63)	247.34 (5.63)	247.34 (5.63)	247.09 (5.63)
	$\overline{m}$	10.27 (1.84)	8.65 (1.68)	8.62 (1.65)	7.33 (1.53)	5.45 (1.72)
1000	$\mathbb{L}_2$	$8.40 \cdot 10^{-3}$	$12.58 \cdot 10^{-3}$	$12.67 \cdot 10^{-3}$	$17.22 \cdot 10^{-3}$	$21.64 \cdot 10^{-3}$
		$(2.77 \cdot 10^{-3})$	$(3.31 \cdot 10^{-3})$	$(3.30 \cdot 10^{-3})$	$(4.27 \cdot 10^{-3})$	$(6.03 \cdot 10^{-3})$
	$\overline{R}$	498.53 (8.00)	497.53 (8.00)	497.53 (8.00)	497.53 (8.00)	497.41 (8.01)
	$\overline{m}$	12.77 (3.20)	10.00 (2.33)	9.92 (2.27)	8.33 (0.47)	7.65 (0.65)
5000	$\mathbb{L}_2$	$3.45 \cdot 10^{-3}$	$4.65 \cdot 10^{-3}$	$4.78 \cdot 10^{-3}$	$12.66 \cdot 10^{-3}$	$21.79 \cdot 10^{-3}$
		$(1.08 \cdot 10^{-3})$	$(1.20 \cdot 10^{-3})$	$(1.11 \cdot 10^{-3})$	$(0.94 \cdot 10^{-3})$	$(203.45 \cdot 10^{-3})$
	$\overline{R}$	2498.90 (17.80)	2497.90 (17.80)	2497.90 (17.80)	2497.90 (17.80)	2497.7 (17.80)
	$\overline{m}$	22.19 (4.44)	18.61 (4.00)	17.91 (3.39)	9.04 (0.45)	8.72 (1.39)

Table 2: Simulation results for τ following a $\mathcal{G}(2, \frac{1}{2})$ distribution under the dead-zone assumption ($\eta = 1$). $\mathbb{L}_2$: mean square errors, $\overline{R}$: mean of the number of observations, $\overline{m}$: mean of the selected dimensions. All standard deviations are given in parenthesis.

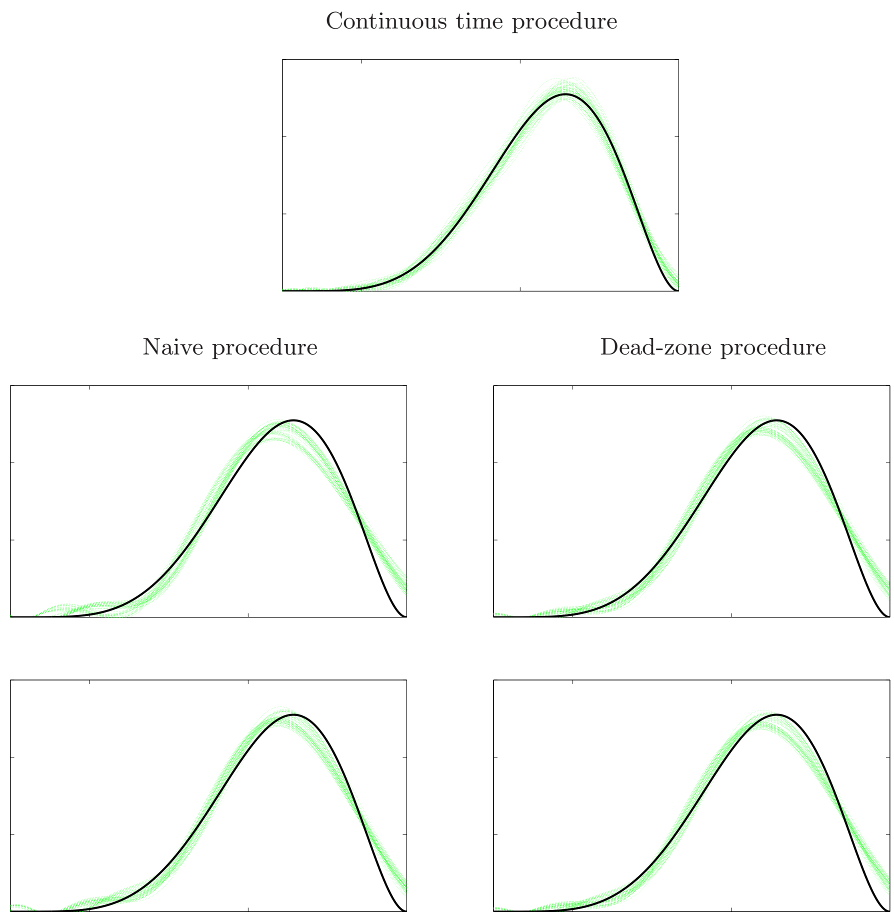

Figure 3: Estimation of τ, following a shifted ($\eta = 1$) Beta distribution $5 \times \mathcal{B}(6, 3)$, for $T = 500$. The estimator based on the continuous observation (first line), $\Delta = 0.1$ (second line) and for $\Delta = 0.01$ (third line), with the naive method (first column) and the dead-zone method (second column). True density τ in black and 50 estimated curves in green.

In the remaining of this section, we discuss how assumptions (A2) and (A4) might be relaxed.

Assumption (A2) is not necessary since it is established in Lindvall (1992) that under (A1) and for large enough T the process R has stationary increments. Then, by removing the first observations, the procedures of Section 3 would have the same properties. Indeed, in the numerical Section all simulated trajectories start from $T_0 = 0$ ((A2) is not satisfied) and the performances of the estimators are consistent with the theoretical results. However, from a theoretical viewpoint, removing assumption (A2) is not straightforward, elements on how one should proceed are given in Duval (2013b).

Removing assumption (A4) is difficult. In the general case, under (A1) and (A2), we may prove that the common density of the observations $(\widehat{D}_j^{\Delta})$ is

$$f_{\Delta}(x) := \Big(\sum_{r=1}^{\infty} \tau^{*r} \frac{\int_0^{\Delta} \tau_0 * \tau^{*r-1}(u) - \tau_0 * \tau^{*r}(u) du}{\int_0^{\Delta} \tau_0(u) du}\Big) * g_{\Delta} * g_{\Delta}(-.)\ (x), \quad \forall x \in \mathbb{R}, \tag{17}$$

where $*$ denotes the convolution product and g_{Δ} is the general density of $F_{\widehat{T}_i^{\Delta} - \Delta}$. The issue remains that (17) is a nonlinear transformation of τ where the transformation itself depends on the knowledge of $\tau \mathbb{1}_{[0,\Delta]}$. Even if we knew $\tau \mathbb{1}_{[0,\Delta]}$ or had access to an estimator, inverting (17) is a difficult problem similar to decompounding (see *e.g.* van Es *et al.* (2007), Duval (2013a,2013b) or Comte *et al.* (2014)). The dead-zone case only partially solves the estimation problem for renewal processes. But, it illustrates that in deconvolution problem, when the Fourier transform of the noise has isolated zero, if Fourier methods become technically difficult, the Laguerre procedure remains easy to implement.

Finally, note that in both continuous and discrete observation schemes, our procedures can be immediately adapted to the case where one observes a renewal reward process X with marks having an unknown distribution that either admits a density with respect to the Lebesgue measure or is positive. Indeed, this last assumption ensures that almost surely if $X_t \neq X_s$, then $R_t \neq R_s$, for all (t, s), consequently all the jumps of R are detected. The estimation of the density of the marks from the discrete observation of X has been studied in Duval (2013b).

6 Proofs

6.1 Proof of Theorem 2.1

Recall that τ_m denotes the orthonormal projection of τ on S_m. By Pythagoras Theorem we have

$$\|\widehat{\tau}_m - \tau\|^2 = \|\tau - \tau_m\|^2 + \sum_{k=0}^{m-1} (\widehat{a}_k - a_k)^2.$$

Taking expectation and decomposing on the possible values of R_T, we are left to control

$$\mathbb{E}[(\widehat{a}_k - a_k)^2] = \mathbb{E}\Big[\sum_{\ell=0}^{\infty} \mathbb{1}_{R_T=\ell} \Big(\frac{1}{\ell} \sum_{i=1}^{\ell} (\varphi_k(D_i) - \langle \varphi_k, \tau \rangle)\Big)^2\Big].$$

As we adopt the convention $\frac{1}{0}\sum_{i=1}^{0} = 0$, the first sum starts at $\ell = 1$. Consider the centered empirical process $\nu_\ell(t) = \frac{1}{\ell}\sum_{i=1}^{\ell} \big(t(D_i) - \langle t, \tau\rangle\big)$, $t \in \mathbb{L}^2(\mathbb{R}^+)$. Then, we can show that

$$\sup_{\|t\|=1,\ t\in S_m} \big(\nu_\ell(t)\big)^2 = \sum_{k=0}^{m-1} \nu_\ell^2(\varphi_k). \tag{18}$$

Indeed, we have using the Cauchy Schwarz inequality that

$$\sup_{\|t\|=1,\ t\in S_m}\big(\nu_\ell(t)\big)^2 = \sup_{(a_k(t))\in\mathbb{R}^m,\ \sum_{k=0}^{m-1}a_k(t)^2=1}\Big(\sum_{k=0}^{m-1}a_k(t)\nu_\ell(\varphi_k)\Big)^2$$

$$\leq \sup_{(a_k(t))\in\mathbb{R}^m,\ \sum_{k=0}^{m-1}a_k(t)^2=1}\Big(\sum_{k=0}^{m-1}a_k(t)^2\Big)\Big(\sum_{k=0}^{m-1}\nu_\ell^2(\varphi_k)\Big) = \sum_{k=0}^{m-1}\nu_\ell^2(\varphi_k).$$

Moreover, if we consider the coefficients

$$a_k(t) := \frac{\nu_\ell(\varphi_k)}{\sqrt{\sum_{k=0}^{m-1}\nu_\ell^2(\varphi_k)}}, \quad k=0,\dots,m-1$$

the former inequality is an equality and (18) is proved. It follows that

$$\sum_{k=0}^{m-1}\mathbb{E}[(\widehat{a}_k-a_k)^2] = \mathbb{E}\Big[\sum_{\ell=1}^{\infty}\mathbb{1}_{R_T=\ell}\sup_{\|t\|=1,\ t\in S_m}\nu_\ell^2(t)\Big]$$

$$\leq \mathbb{E}\Big[\sum_{\ell=1}^{\infty}\mathbb{1}_{R_T=\ell}\big(\sup_{\|t\|=1,\ t\in S_m}\nu_\ell^2(t)-2(1+2\varepsilon_\ell)H_\ell^2)\big)_+\Big]+\mathbb{E}\Big[\sum_{\ell=1}^{\infty}\mathbb{1}_{R_T=\ell}2(1+2\varepsilon_\ell)H_\ell^2\Big]$$

$$\leq \sum_{\ell=1}^{\infty}\mathbb{P}(R_T=\ell)^{1/2}\mathbb{E}^{1/2}\big[\big(\sup_{\|t\|=1,\ t\in S_m}\nu_\ell^2(t)-2(1+2\varepsilon_\ell)H_\ell^2)\big)_+^2\big]+\mathbb{E}\Big[\sum_{\ell=1}^{\infty}\mathbb{1}_{R_T=\ell}2(1+2\varepsilon_\ell)H_\ell^2\Big],$$

for any positive constants ε_ℓ and H_ℓ. We want to apply Lemma 6.2 (see the Appendix section) with $\mathcal{F}=\{t\in S_m, \|t\|=1\}$, and a classical density argument. To apply Lemma 6.2 we compute b, v and H_ℓ. If $t\in S_m$ such that $\|t\|=1$ we have $\|t\|_\infty \leq \sqrt{2}\sum_{k=0}^{m-1}|a_k(t)| \leq \sqrt{2m} := b$, by the Cauchy Schwarz inequality. Next, we note that

$$\sup_{\|t\|=1,\ t\in S_m}\mathbb{E}\big[t(D_1)^2\big] \leq \sup_{\|t\|=1,\ t\in S_m}\|t\|_\infty\int_0^\infty |t(x)\tau(x)|dx \leq \sqrt{2m}\|\tau\| := v.$$

Finally, using (18) and that ν_ℓ is centered we get $\mathbb{E}\big[\sup_{\|t\|=1,\ t\in S_m}\big|\nu_\ell(t)\big|^2\big] \leq \frac{2m}{\ell} := H_\ell^2$. To summarize we have

$$H_\ell = \sqrt{\frac{2m}{\ell}}, \quad v=\sqrt{2m}\|\tau\| \quad \text{and} \quad b=\sqrt{2m}. \tag{19}$$

It follows from Lemma 6.2 (see the Appendix section), with parameters (19) and $\varepsilon_\ell=\frac{1}{2}$, that

$$\sum_{k=0}^{m-1}\mathbb{E}[(\widehat{a}_k-a_k)^2] \leq \sum_{\ell=1}^{\infty}\mathbb{P}(R_T=\ell)^{\frac{1}{2}}\Big(6\Big(\frac{\sqrt{8m}\|\tau\|}{\ell\kappa'}\Big)^2\exp\Big(-\frac{\kappa'\sqrt{m}}{\sqrt{2}\|\tau\|}\Big)+36\Big(\frac{2\sqrt{2m}}{\ell\kappa'}\Big)^4\exp\Big(-\frac{\sqrt{\ell}\kappa'}{2}\Big)\Big)^{\frac{1}{2}}$$

$$+\mathbb{E}\Big[\sum_{\ell=1}^{\infty}\mathbb{1}_{R_T=\ell}\frac{8m}{\ell}\Big]$$

where κ' is a universal constant. From the Cauchy Schwarz inequality and $\sqrt{a+b}\leq\sqrt{a}+\sqrt{b}$, we get

$$\sum_{k=0}^{m-1}\mathbb{E}[(\widehat{a}_k-a_k)^2] \leq \sqrt{\sum_{\ell=1}^{\infty}\mathbb{P}(R_T=\ell)}\sqrt{\sum_{\ell=1}^{\infty}6\Big(\frac{\sqrt{8}\|\tau\|}{\ell\kappa'}\Big)^2 m\exp\Big(-\frac{\kappa'}{2\sqrt{2}\|\tau\|}\sqrt{m}\Big)}$$

$$+6\sqrt{\sum_{\ell=1}^{\infty}\mathbb{P}(R_T=\ell)\Big(\frac{2\sqrt{2m}}{\ell\kappa'}\Big)^4}\Big(\sum_{\ell=1}^{\infty}e^{-\frac{\sqrt{\ell}\kappa'}{2}}\Big)^{1/2}+8m\mathbb{E}\Big[\frac{\mathbb{1}_{R_T\geq 1}}{R_T}\Big]$$

$$=\mathbf{C}_1\|\tau\|^3\exp\Big(-\frac{\kappa'}{4\sqrt{2}\|\tau\|}\sqrt{m}\Big)+\mathbf{C}_2 m\sqrt{\mathbb{E}\Big[\frac{\mathbb{1}_{R_T\geq 1}}{R_T^4}\Big]}+8m\mathbb{E}\Big[\frac{\mathbb{1}_{R_T\geq 1}}{R_T}\Big],$$

where we set

$$\mathbf{C}_1 = \frac{2^7\sqrt{3}}{\kappa'^3 e^2}\sqrt{\sum_{\ell=1}^{\infty}\frac{1}{\ell^2}} \quad \text{and} \quad \mathbf{C}_2 = \frac{48}{(\kappa')^2}\Big(\sum_{\ell=1}^{\infty} e^{-\frac{\sqrt{\ell}\kappa'}{2}}\Big)^{1/2}, \tag{20}$$

note that we used that $xe^{-\theta\sqrt{x}} \leq \left(\frac{2}{e\theta}\right)^2$, for $x, \theta > 0$, thus $xe^{-2\theta\sqrt{x}} \leq \left(\frac{2}{e\theta}\right)^2 e^{-\theta\sqrt{x}}$. Gathering all the terms completes the proof. □

6.2 Proof of Proposition 2.1

Recall that $T_\ell = \sum_{j=1}^{\ell} D_j$. Using the definition of R_T it is straightforward to establish the following

$$\frac{T_{R_T}}{R_T}\mathbb{1}_{R_T\geq 1} \leq \frac{T}{R_T}\mathbb{1}_{R_T\geq 1} \leq \frac{T_{R_T+1}}{R_T}\mathbb{1}_{R_T\geq 1}, \quad \forall T > 0. \tag{21}$$

We have the decomposition

$$\mathbb{E}\Big[\Big(\frac{T_{R_T}}{R_T}\Big)^{\alpha}\mathbb{1}_{R_T\geq 1}\Big] = \mathbb{E}\Big[\sum_{\ell=1}^{\infty}\Big(\frac{T_\ell}{\ell}\Big)^{\alpha}\mathbb{1}_{R_T=\ell}\Big].$$

Introduce the event

$$\widetilde{\Omega}_\ell = \Big\{\Big|\frac{T_\ell}{\ell} - \mu\Big| \leq \frac{\mu}{2}\Big\}.$$

First let $\alpha > 0$, it is easy to get

$$\mathbb{E}\Big[\Big(\frac{T_{R_T}}{R_T}\Big)^{\alpha}\mathbb{1}_{R_T\geq 1}\Big] \geq \mathbb{E}\Big[\sum_{\ell=1}^{\infty}\Big(\frac{\mu}{2}\Big)^{\alpha}\mathbb{1}_{R_T=\ell}\Big] = \Big(\frac{\mu}{2}\Big)^{\alpha}\mathbb{P}(R_T \geq 1) \geq \mathsf{a}\Big(\frac{\mu}{2}\Big)^{\alpha}. \tag{22}$$

Moreover, under (A3) we apply the Bernstein inequality (see Corollary 2.10 in Massart [25]) to get

$$\mathbb{P}(\widetilde{\Omega}_\ell^c) \leq 2\exp\Big(-\frac{\ell\mu^2}{8(\sigma^2 + c\frac{\mu}{2})}\Big). \tag{23}$$

We derive, using that $\frac{\ell+1}{\ell} \leq 2$, $\forall \ell \geq 1$ and that $\alpha > 0$,

$$\begin{aligned}\mathbb{E}\Big[\Big(\frac{T_{R_T+1}}{R_T}\Big)^{\alpha}\mathbb{1}_{R_T\geq 1}\Big] &\leq 2^{\alpha}\mathbb{E}\Big[\Big(\frac{T_{R_T+1}}{R_T+1}\Big)^{\alpha}\Big] \leq \mathbb{E}\Big[\sum_{\ell=1}^{\infty}(3\mu)^{\alpha}\mathbb{1}_{R_T=\ell}\Big] + \mathbb{E}\Big[\sum_{\ell=1}^{\infty}\Big(\frac{T_\ell}{\ell}\Big)^{\alpha}\mathbb{1}_{R_T=\ell}\mathbb{1}_{\widetilde{\Omega}_\ell^c}\Big] \\ &\leq (3\mu)^{\alpha} + \sum_{\ell=1}^{\infty}\sqrt{\mathbb{E}\Big[\Big(\frac{T_\ell}{\ell}\Big)^{2\alpha}\Big]\mathbb{E}\Big[\mathbb{1}_{\widetilde{\Omega}_\ell^c}\mathbb{1}_{R_T=\ell}\Big]}.\end{aligned}$$

If $\alpha \geq 1/2$, $x \to x^{2\alpha}$ is convex, together with (A3), (23) and the Cauchy Schwarz inequality we obtain

$$\begin{aligned}\mathbb{E}\Big[\Big(\frac{T_{R_T+1}}{R_T}\Big)^{\alpha}\mathbb{1}_{R_T\geq 1}\Big] &\leq (3\mu)^{\alpha} + \sqrt{\frac{\lceil 2\alpha\rceil!\sigma^2 c^{\lceil 2\alpha\rceil-2}}{2}}\sum_{\ell=1}^{\infty}(\mathbb{P}(\widetilde{\Omega}_\ell^c)\mathbb{P}(R_T=\ell))^{\frac{1}{4}} \\ &\leq (3\mu)^{\alpha} + \sqrt{\frac{\lceil 2\alpha\rceil!\sigma^2 c^{\lceil 2\alpha\rceil-2}}{2}}\sum_{\ell=1}^{\infty}2^{\frac{1}{4}}\exp\Big(-\frac{\ell\mu^2}{32(\sigma^2+c\frac{\mu}{2})}\Big) \\ &\leq (3\mu)^{\alpha} + \sqrt{\frac{\lceil 2\alpha\rceil!\sigma^2 c^{\lceil 2\alpha\rceil-2}}{\sqrt{2}}}\Big(1 - e^{-\frac{\mu^2}{32(\sigma^2+c\frac{\mu}{2})}}\Big)^{-1}. \end{aligned} \tag{24}$$

Now if $0 < \alpha < \frac{1}{2}$, $x \to x^{2\alpha}$ is concave, using the Jensen inequality and similar arguments as above, we get

$$\mathbb{E}\Big[\Big(\frac{T_{R_T+1}}{R_T}\Big)^{\alpha}\mathbb{1}_{R_T\geq 1}\Big] \leq (3\mu)^{\alpha} + \mathbb{E}[D_1]^{\alpha}\Big(1 - e^{-\frac{\mu^2}{32(\sigma^2+c\frac{\mu}{2})}}\Big)^{-1}. \tag{25}$$

Finally, gathering equations (24) , (25) and (22) into (21) and taking expectation provides the following under (A3) and for $\alpha > 0$

$$C_2 T^{-\alpha} \leq \mathbb{E}\Big[\frac{\mathbb{1}_{R_T \geq 1}}{R_T^\alpha}\Big] \leq C_1 T^{-\alpha},$$

where C_2 is defined in (22) and C_1 in (24) if $\alpha \geq 1/2$ or in (25) if $\alpha \in (0, \frac{1}{2})$. This completes the proof. □

6.3 Proof of Theorem 2.2

6.3.1 Proof of Theorem 2.2

First, observe that

$$\hat{m} = \arg\min_{m \in \mathcal{M}_T} \big(-\|\widehat{\tau}_m\|^2 + \widehat{\mathrm{pen}}(m)\big) = \arg\min_{m \in \mathcal{M}_T} \big(\|\tau - \widehat{\tau}_m\|^2 + \widehat{\mathrm{pen}}(m)\big)$$

Consider the contrast $\gamma_{R_T}(t) = \|t\|^2 - (2/R_T)\sum_{i=1}^{R_T} t(D_i)$. It is easily verified that, $\widehat{\tau}_m = \arg\min_{t \in S_m} \gamma_T(t)$. Moreover, we note that

$$\gamma_{R_T}(t) - \gamma_{R_T}(s) = \|t - \tau\|^2 - \|s - \tau\|^2 + 2\langle t - s, \tau\rangle - \frac{2}{R_T}\sum_{i=1}^{R_T}(t - s)(D_i). \tag{26}$$

Then, by definition of $\widehat{m}$, we have $\gamma_T(\widehat{\tau}_{\widehat{m}}) + \widehat{\mathrm{pen}}(\widehat{m}) \leq \gamma_T(\tau_m) + \widehat{\mathrm{pen}}(m)$. This with (26) implies

$$\|\widehat{\tau}_{\widehat{m}} - \tau\|^2 \leq \|\tau - \tau_m\|^2 + \widehat{\mathrm{pen}}(m) + 2\nu_{R_T}(\widehat{\tau}_{\widehat{m}} - \tau_m) - \widehat{\mathrm{pen}}(\widehat{m}), \tag{27}$$

where

$$\nu_{R_T}(t) = \frac{1}{R_T}\sum_{i=1}^{R_T}\big(t(D_i) - \langle t, \tau\rangle\big).$$

Using that ν_{R_T} is a linear form and the inequality $2xy \leq \frac{1}{4}x^2 + 4y^2$ we get

$$\begin{aligned} 2\nu_{R_T}(\widehat{\tau}_{\widehat{m}} - \tau_m) &\leq \frac{1}{4}\|\widehat{\tau}_{\widehat{m}} - \tau_m\|^2 + 4\sup_{t \in S_{\widehat{m}\vee m}, \|t\|=1} \nu_{R_T}(t)^2 \\ &\leq \frac{1}{2}\|\widehat{\tau}_{\widehat{m}} - \tau\|^2 + \frac{1}{2}\|\tau_m - \tau\|^2 + 4\sup_{t \in S_{\widehat{m}\vee m}, \|t\|=1} \nu_{R_T}(t)^2. \end{aligned}$$

Plugging this in (27) and gathering the terms, lead to

$$\frac{1}{2}\|\widehat{\tau}_{\widehat{m}} - \tau\|^2 \leq \frac{3}{2}\|\tau - \tau_m\|^2 + \widehat{\mathrm{pen}}(m) + 4\sup_{t \in S_{\widehat{m}\vee m}, \|t\|=1} \nu_{R_T}(t)^2 - \widehat{\mathrm{pen}}(\widehat{m}).$$

We introduce the following Lemma (see the proof in Section 6.3.2):

Lemma 6.1. *Under the Assumptions of Theorem 2.2, let*

$$p_{R_T}(m) = 2(1 + 2\mathfrak{c}\log(1 + R_T))\frac{2m}{R_T}\mathbb{1}_{R_T \geq 1}. \tag{28}$$

For $\mathfrak{c} \geq \max(1/(\sqrt{2}\kappa'), 2/(\log(2)\kappa')^2)$, where κ' is defined in Corollary 2 of Birgé and Massart [5], we have, for $T \geq e^{6\|\tau\|}$,

$$\mathbb{E}\left[\left(\sup_{t \in S_{\widehat{m}\vee m}, \|t\|=1} \nu_{R_T}(t)^2 - p_{R_T}(\widehat{m} \vee m)\right)_+\right] \leq 2c'\mathbb{E}^{1/2}\left[\frac{T^4\mathbb{1}_{R_T \geq 1}}{R_T^6}\right]$$

with c' given in (33).

We have

$$\frac{1}{2}\|\widehat{\tau}_{\widehat{m}}-\tau\|^2 \leq \frac{3}{2}\|\tau-\tau_m\|^2+4\Big(\sup_{t\in S_{\widehat{m}\vee m},\|t\|=1}\nu_{R_T}(t)^2-p_{R_T}(m\vee\widehat{m})\Big)_+ + \widehat{\mathrm{pen}}(m)+4p_{R_T}(m\vee\widehat{m})-\widehat{\mathrm{pen}}(\widehat{m})$$

where p_{R_T} is defined in (28). Using that $4p_{R_T}(m\vee\widehat{m})\leq 4p_{R_T}(m)+4p_{R_T}(\widehat{m})$ and $\widehat{\mathrm{pen}}(m')=4p_{R_T}(m')$, $\forall m'$, we get

$$\frac{1}{2}\|\widehat{\tau}_{\widehat{m}}-\tau\|^2 \leq \frac{3}{2}\|\tau-\tau_m\|^2+4\Big(\sup_{t\in S_{\widehat{m}\vee m},\|t\|=1}\nu_{R_T}(t)^2-p_{R_T}(m\vee\widehat{m})\Big)_+ +2\widehat{\mathrm{pen}}(m) \tag{29}$$

Taking expectation in (29) together with Lemma 6.1, we derive $\forall m\in\mathcal{M}_T$

$$\mathbb{E}[\|\widehat{\tau}_{\widehat{m}}-\tau\|^2]\leq 3\|\tau-\tau_m\|^2+4\mathbb{E}[\widehat{\mathrm{pen}}(m)]+16c'\mathbb{E}^{1/2}\left[\frac{T^4\mathbb{1}_{R_T\geq 1}}{R_T^6}\right].$$

This implies the result given in Theorem 2.2. □

6.3.2 Proof of Lemma 6.1

First, we use that

$$\begin{aligned}\mathbb{E}\Big[\Big(\sup_{t\in S_m,\|t\|=1}\nu_{R_T}(t)^2-p_{R_T}(m)\Big)_+\Big] &= \sum_{\ell=0}^{\infty}\mathbb{E}\left[\Big(\sup_{t\in S_m,\|t\|=1}\nu_\ell(t)^2-p_\ell(m)\Big)_+\mathbb{1}_{R_T=\ell}\right]\\ &\leq \sum_{\ell=0}^{\infty}\Big(\mathbb{E}\Big[\Big(\sup_{t\in S_m,\|t\|=1}\nu_\ell(t)^2-p_\ell(m)\Big)_+^2\Big]\mathbb{P}(R_T=\ell)\Big)^{\frac{1}{2}},\end{aligned} \tag{30}$$

where $\nu_\ell(t)=\frac{1}{\ell}\sum_{j=1}^{\ell}\big(t(D_j)-\langle t,\tau\rangle\big)$ and $p_\ell(m)=2(1+2\varepsilon_\ell)H_\ell^2$, with the convention $\nu_0(t)=0$, $\forall t$ and $p_0(m)=0$, $\forall m$, so that the previous sum starts at $\ell=1$. We bound the expectation in (30) applying Lemma 6.2 (see the Appendix section) as in the proof of Theorem 2.1 with b, v and H_ℓ given by (19). Now we take $\varepsilon_\ell=\mathfrak{c}\log(1+\ell)$. Denote by $X=\Big(\sup_{t\in S_m,\|t\|=1}\nu_\ell(t)^2-2(1+\varepsilon_\ell)H_\ell^2\Big)_+$, we obtain

$$\begin{aligned}\mathbb{E}\big[X^2\big] &\leq 6\Big(\frac{\sqrt{8m}\|\tau\|}{\ell\kappa'}\Big)^2\exp\Big(-\frac{\kappa'\mathfrak{c}\sqrt{2}}{\|\tau\|}\sqrt{m}\log(1+\ell)\Big)+36\Big(\frac{2\sqrt{2m}}{\ell\kappa'}\Big)^4\exp\Big(-\frac{\kappa'\sqrt{\mathfrak{c}\ell\log(1+\ell)}}{\sqrt{2}}\Big)\\ &\leq \frac{2^4 3\|\tau\|^2}{(\kappa')^2}\frac{m}{\ell^{\sqrt{m}/\|\tau\|+2}}+\frac{2^8 3^2 m^2}{(\kappa')^4\ell^4}\exp\big(-\sqrt{\ell}\big),\end{aligned} \tag{31}$$

for $\mathfrak{c}\geq(1/(\sqrt{2}\kappa'))\vee 2/(\log(2)\kappa')^2)$. Plugging (31) into (30), together with the inequality $\sqrt{a+b}\leq\sqrt{a}+\sqrt{b}$ and using that $\ell^4\exp(-\sqrt{\ell})\leq c_0=(8/e)^8$, leads to

$$\begin{aligned}&\mathbb{E}\Big[\Big(\sup_{t\in S_m,\|t\|=1}\nu_{R_T}(t)^2-p(m)\Big)_+\Big]\\ &\leq \frac{\sqrt{2^4 3m}\|\tau\|}{\kappa'}\sum_{\ell=1}^{\infty}\sqrt{\frac{\mathbb{P}(R_T=\ell)}{\ell^{2+\sqrt{m}/\|\tau\|}}}+\frac{2^4 3m\sqrt{c_0}}{(\kappa')^2}\sum_{\ell=1}^{\infty}\Big(\frac{\mathbb{P}(R_T=\ell)}{\ell^8}\Big)^{1/2}.\\ &\leq \frac{4\sqrt{3m}\|\tau\|}{\kappa'}\sqrt{\sum_{\ell=1}^{\infty}\frac{1}{\ell^2}}\sqrt{\mathbb{E}\left[\frac{\mathbb{1}_{R_T\geq1}}{(R_T)^{\sqrt{m}/\|\tau\|}}\right]}+\frac{2^4 3m\sqrt{c_0}}{(\kappa')^2}\sqrt{\sum_{\ell=1}^{\infty}\frac{1}{\ell^2}}\sqrt{\mathbb{E}\left[\frac{\mathbb{1}_{R_T\geq1}}{R_T^6}\right]},\end{aligned} \tag{32}$$

where the last inequality follows from the Cauchy Schwarz inequality. To conclude, we write that

$$\begin{aligned}\mathbb{E}\left[\Big(\sup_{t\in S_{\widehat{m}\vee m},\|t\|=1}\nu_{R_T}(t)^2-p_{R_T}(\widehat{m}\vee m)\Big)_+\right] &\leq \sum_{m'\in\mathcal{M}_T}\mathbb{E}\left[\Big(\sup_{t\in S_{m'\vee m},\|t\|=1}\nu_{R_T}(t)^2-p_{R_T}(m\vee m')\Big)_+\right]\\ &\leq c'\sum_{m'\in\mathcal{M}_T}\left(\sqrt{m\vee m'}\sqrt{\mathbb{E}\left[\frac{\mathbb{1}_{R_T\geq1}}{(R_T)^{(\sqrt{m'}\vee\sqrt{m})/\|\tau\|}}\right]}+m'\vee m\sqrt{\mathbb{E}\left[\frac{\mathbb{1}_{R_T\geq1}}{R_T^6}\right]}\right),\end{aligned}$$

for c' a constant, $c' \geq \max(C_1, C_2)$, where C_1 and C_2 can be derived from (32)

$$C_1 = \frac{4\sqrt{3}\|\tau\|}{\kappa'}\sqrt{\sum_{\ell=1}^{\infty}\frac{1}{\ell^2}}, \quad C_2 = \frac{2^4 3\sqrt{c_0}}{(\kappa')^2}\sqrt{\sum_{\ell=1}^{\infty}\frac{1}{\ell^2}}. \tag{33}$$

Consequently, using that $\log^2(T) \leq m \leq T$ together with the Cauchy Schwarz inequality, we get

$$\mathbb{E}\left[\left(\sup_{t\in S_{\widehat{m}\vee m}, \|t\|=1} \nu_{R_T}(t)^2 - p_{R_T}(m)\right)_+\right] \leq c'\left[T\sum_{m=\lfloor \log^2(T)\rfloor}^{T}\sqrt{\mathbb{E}\Big[\frac{\mathbb{1}_{R_T\geq 1}}{R_T^{\sqrt{m}/\|\tau\|}}\Big]} + T^2\sqrt{\mathbb{E}\Big[\frac{\mathbb{1}_{R_T\geq 1}}{R_T^6}\Big]}\right]. \tag{34}$$

Now, for $T \geq \exp(6\|\tau\|)$, we have $\sqrt{m}/\|\tau\| \geq \log(T)/\|\tau\| \geq 6$ and

$$\mathbb{E}\Big[\frac{\mathbb{1}_{R_T\geq 1}}{R_T^{\sqrt{m}}}\Big] \leq \mathbb{E}\Big[\frac{\mathbb{1}_{R_T\geq 1}}{R_T^6}\Big]. \tag{35}$$

Therefore, plugging (35) in equation (34) implies the result of Lemma 6.1. □

6.4 Proof of Corollary 2.2

It follows from the Cauchy Schwarz inequality that

$$\mathbb{E}\Big[\frac{\log(1+R_T)}{R_T}\mathbb{1}_{R_T\geq 1}\Big] \leq \sqrt{\mathbb{E}\Big[\frac{\mathbb{1}_{R_T\geq 1}}{R_T^2}\Big]}\sqrt{\mathbb{E}\big[\big(\log(1+R_T)\big)^2\big]}.$$

The function $x \to \big(\log(1+x)\big)^2$ is concave for $xe \geq 1$. Then, decomposing on the events $\{R_T \leq 1\}$ and $\{R_T \geq 2\}$ and applying the Jensen inequality leads to,

$$\sqrt{\mathbb{E}\big[\big(\log(1+R_T)\big)^2\big]} \leq \sqrt{\big(\log(2)\big)^2 + \big(\log(1+\mathbb{E}[R_T])\big)^2} \leq \log(2) + \log(1+\mathbb{E}[R_T]).$$

Next, using the inequality (see Grimmett and Stirzaker (2001) p. 420)

$$\mathbb{E}[R_T] \leq \frac{T}{\mu_1} + \frac{1-\mu_1}{\mu_1}$$

where $\mu_1 = \mathbb{E}[D_1 \wedge 1] > 0$, leads to

$$\mathbb{E}\big[\log(1+R_T)\big] \leq \Big|\log\Big(\frac{T+1}{\mu_1}\Big)\Big|.$$

Finally, Inequality (4) of Proposition 2.1 with $\alpha = 2$ gives

$$\mathbb{E}\Big[\frac{\log(1+R_T)}{R_T}\mathbb{1}_{R_T\geq 1}\Big] \leq \frac{\sqrt{C_2}}{T}\big(C_3 + \log(T+1)\big),$$

where C_1 is defined in Proposition 2.1 and $C_3 = \log(2) + |\log(\mu_1)|$. This completes the proof. □

6.5 Proof of Lemma 3.1

From (6), we derive that for $i \geq 1$, $\widehat{D}_{i+1}^{\Delta} = D_{R_{\widehat{T}_i^{\Delta}}+1} + \Delta\xi_i$ where we set $\xi_i := \frac{1}{\Delta}\big(F_{\widehat{T}_i^{\Delta}-\Delta} - F_{\widehat{T}_{i+1}^{\Delta}-\Delta}\big)$. By the definition of forward times and the variables $(\widehat{T}_i^{\Delta})$, it is straightforward to get that $|\xi_i| \leq 1$. We are left to prove that $(D_{R_{\widehat{T}_i^{\Delta}}+1})$ are i.i.d. with density τ. The independence is due to the renewal property, we prove the density is τ. Let $h : \mathbb{R}^+ \to \mathbb{R}$ be a bounded measurable function, decomposing on the values of $\widehat{T}_i^{\Delta}$, we find that

$$\mathbb{E}\big[h(D_{R_{\widehat{T}_i^{\Delta}}+1})\big] = \sum_{j=1}^{\lfloor T\Delta^{-1}\rfloor}\mathbb{E}\big[h(D_{R_{j\Delta}+1})\big|\widehat{T}_i^{\Delta} = j\Delta\big]\mathbb{P}(\widehat{T}_i^{\Delta} = j\Delta).$$

It is sufficient to show that:

$$\text{for all } k \leq j \text{ the variables } D_{R_{j\Delta}+1} \text{ and } (R_{k\Delta} - R_{(k-1)\Delta}) \text{ are independent} \tag{36}$$

and that

$$D_{R_{j\Delta}+1} \text{ has density } \tau. \tag{37}$$

Indeed, if (36) and (37) hold true, the independence between $D_{R_{j\Delta}+1}$ and $(R_{k\Delta} - R_{(k-1)\Delta})_{k\leq j}$ ensures that $D_{R_{j\Delta}+1}$ is independent of the event $\{\widehat{T}_i^\Delta = j\Delta\}$. This leads to

$$\begin{aligned}
\mathbb{E}\big[h(D_{R_{\widehat{T}_i^\Delta}+1})\big] &= \sum_{j=1}^{\lfloor T\Delta^{-1}\rfloor} \mathbb{E}\big[h(D_{R_{j\Delta}+1})\big]\mathbb{P}(\widehat{T}_i^\Delta = j\Delta) \\
&= \sum_{j=1}^{\lfloor T\Delta^{-1}\rfloor} \int_0^\infty h(y)\tau(y)dy\mathbb{P}(\widehat{T}_i^\Delta = j\Delta) = \mathbb{E}\big[h(D_1)\big].
\end{aligned}$$

Therefore, this implies that $D_{R_{\widehat{T}_i^\Delta}+1}$ has density τ.

Now, we prove (36) and (37). Let $h_1 : \mathbb{R}^+ \to \mathbb{R}$ and $h_2 : \mathbb{N} \to \mathbb{R}$ be bounded measurable functions, and $k \leq j$. We have

$$\begin{aligned}
&\mathbb{E}\big[h_1(D_{R_{j\Delta}+1})h_2(R_{k\Delta} - R_{(k-1)\Delta})\big] \\
&\quad= \sum_{\ell_1=0}^{\infty}\sum_{\ell_2=0}^{\ell_1}\sum_{\ell_3=0}^{\ell_2} h_2(\ell_3)\mathbb{E}\big[h_1(D_{\ell_1+1})\big|R_{j\Delta} = \ell_1, R_{k\Delta} = \ell_2, R_{k\Delta} - R_{(k-1)\Delta} = \ell_3\big] \\
&\qquad\qquad \times \mathbb{P}\big(R_{j\Delta} = \ell_1, R_{k\Delta} = \ell_2, R_{k\Delta} - R_{(k-1)\Delta} = \ell_3\big).
\end{aligned}$$

As $k \leq j$, we have $R_{k\Delta} - R_{(k-1)\Delta} \leq R_{k\Delta} \leq R_{j\Delta}$ *a.s.* and the renewal property ensures that D_{ℓ_1+1} is independent of the event $\{R_{j\Delta} = \ell_1, R_{k\Delta} = \ell_2, R_{k\Delta} - R_{(k-1)\Delta} = \ell_3\}$, $0 \leq \ell_3 \leq \ell_2 \leq \ell_1$, it follows that

$$\begin{aligned}
&\mathbb{E}\big[h_1(D_{R_{j\Delta}+1})h_2(R_{k\Delta} - R_{(k-1)\Delta})\big] \\
&\quad= \mathbb{E}\big[h_1(D_1)\big]\sum_{\ell_3=0}^{\infty} h_2(\ell_3)\sum_{\ell_1=\ell_3}^{\infty}\sum_{\ell_2=\ell_3}^{\ell_1} \mathbb{P}\big(R_{j\Delta} = \ell_1, R_{k\Delta} = \ell_2, R_{k\Delta} - R_{(k-1)\Delta} = \ell_3\big) \\
&\quad= \mathbb{E}\big[h_1(D_1)\big]\sum_{\ell_3=0}^{\infty} h_2(\ell_3)\mathbb{P}\big(R_{j\Delta} \geq R_{k\Delta}, R_{k\Delta} \geq R_{k\Delta} - R_{(k-1)\Delta}, R_{k\Delta} - R_{(k-1)\Delta} = \ell_3\big) \\
&\quad= \mathbb{E}\big[h_1(D_1)\big]\sum_{\ell_3=0}^{\infty} h_2(\ell_3)\mathbb{P}\big(R_{k\Delta} - R_{(k-1)\Delta} = \ell_3\big) = \mathbb{E}\big[h_1(D_1)\big]\mathbb{E}\big[h_2(R_{k\Delta} - R_{(k-1)\Delta})\big]
\end{aligned}$$

where in last line, we use that $k \leq j$, implying that $R_{k\Delta} - R_{(k-1)\Delta} \leq R_{k\Delta} \leq R_{j\Delta}$ *a.s.* The equality implies both (36) and (37). The proof of Lemma 3.1 is now complete. □

6.6 Proof of Proposition 3.1

As in the proof of Theorem 2.1 we have

$$\|\check{\tau}_m - \tau\|^2 = \|\tau - \tau_m\|^2 + \sum_{k=0}^{m-1}(\check{a}_k - a_k)^2.$$

Having an expansion of the coefficients $\check{a}_k$ based on relation (8) leads to

$$\check{a}_k = \frac{1}{N_T}\sum_{i=1}^{N_T}\varphi_k(\widehat{D}_i^\Delta) = \frac{1}{N_T}\sum_{i=1}^{N_T}\varphi_k\big(D_{R_{\widehat{T}_i^\Delta}+1} + \Delta\xi_i\big) = \widetilde{a}_k + \frac{\Delta}{N_T}\sum_{i=1}^{N_T}\varphi_k'(\widetilde{\xi}_i),$$

for some random variables $\widetilde{\xi}_j$ and where

$$\widetilde{a}_k := \frac{1}{N_T}\sum_{i=1}^{N_T}\varphi_k\big(D_{R_{\widehat{T}_i^\Delta}+1}\big).$$

It follows that

$$\sum_{k=0}^{m-1}(\breve{a}_k - a_k)^2 \le 2\sum_{k=0}^{m-1}(\widetilde{a}_k - a_k)^2 + 2\Delta^2\sum_{k=0}^{m-1}\Big(\frac{1}{N_T}\sum_{i=1}^{N_T}|\varphi_k'(\widetilde{\xi}_i)|\Big)^2.$$

Using that $\|\varphi_k\|_\infty \le \sqrt{2}$, $\forall k$ and the relation (see Lemma 5.2 in Comte and Dion (2016))

$$\varphi_k' = -\varphi_k - 2\sum_{\ell=0}^{k-1}\varphi_\ell \tag{38}$$

we get $\|\varphi_k'\|_\infty \le \sqrt{2}(1+2k)$. This leads to

$$\begin{aligned}\sum_{k=0}^{m-1}(\breve{a}_k - a_k)^2 &\le 2\sum_{k=0}^{m-1}(\widetilde{a}_k - a_k)^2 + 2\Delta^2\sum_{k=0}^{m-1}2(1+2k)^2\\ &= 2\sum_{k=0}^{m-1}(\widetilde{a}_k - a_k)^2 + \Delta^2\frac{m(4m^2-1)}{3}.\end{aligned}$$

Taking expectation and thanks to Lemma 3.1 the first term can be treated similarly as in the proof of Theorem 2.1 replacing R_T with N_T. We derive Proposition 3.1. □

6.7 Proof of Theorem 3.1

The proof of Theorem 3.1 follows the line of the proof of Theorem 2.2 with $\nu_{R_T}(t)$ replaced by $\breve{\nu}_{N_T}(t)$ where

$$\breve{\nu}_{N_T}(t) = \frac{1}{N_T}\sum_{i=1}^{N_T}(t(\widehat{D}_i^\Delta) - \langle\tau, t\rangle).$$

We have

$$\sup_{t\in S_{\breve{m}\vee m}, \|t\|=1}[\breve{\nu}_{N_T}(t)]^2 \le 2\sup_{t\in S_{\breve{m}\vee m}, \|t\|=1}[\nu_{N_T}(t)]^2 + 2\sup_{t\in S_{\breve{m}\vee m}, \|t\|=1}[r\breve{e}s_T(t)]^2$$

where

$$r\breve{e}s_T(t) = \frac{1}{N_T}\sum_{i=1}^{N_T}(t(\widehat{D}_i^\Delta) - t(D_i)).$$

It follows from the proof of Proposition 3.1 that

$$\sup_{t\in S_{\breve{m}\vee m}, \|t\|=1}[r\breve{e}s_T(t)]^2 \le \frac{4}{3}m^3\Delta^2.$$

Then, let $p_{N_T}(m)$ defined in (28) and $\breve{p}_{N_T}(m) = (8/3)\Delta^2 m^3$. We get

$$\begin{aligned}\frac{1}{2}\|\breve{\tau}_{\breve{m}} - \tau\|^2 \quad &\le \quad \frac{3}{2}\|\tau - \tau_m\|^2 + \text{p}\breve{\text{e}}\text{n}(m) + 8\Big(\sup_{t\in S_{\hat{m}\vee m}, \|t\|=1}\nu_{N_T}(t)^2 - p_{N_T}(m\vee\hat{m})\Big)_+\\ &\quad +8\Big(\sup_{t\in S_{\breve{m}\vee m}, \|t\|=1}[r\breve{e}s_T(t)]^2 - \breve{p}_{N_T}(m\vee\hat{m})\Big)_+ + 8p_{N_T}(m\vee\breve{m})\\ &\quad +8\breve{p}_{N_T}(m\vee\breve{m}) - \text{p}\breve{\text{e}}\text{n}(\breve{m})\\ &\le \quad \frac{3}{2}\|\tau - \tau_m\|^2 + \text{pen}(m) + 8\Big(\sup_{t\in S_{\hat{m}\vee m}, \|t\|=1}\nu_{N_T}(t)^2 - p_{N_T}(m\vee\hat{m})\Big)_+\\ &\quad +8p_{N_T}(m\vee\breve{m}) + 8\breve{p}_{N_T}(m\vee\breve{m}) - \text{p}\breve{\text{e}}\text{n}(\breve{m}).\end{aligned}$$

Now we choose $\check{\text{pen}}(m) = 8p_{N_T}(m) + 8\check{p}_{T,2}(m)$ so that

$$8p_{N_T}(m \vee \check{m}) + 8\check{p}_{N_T}(m \vee \check{m}) - \check{\text{pen}}(\check{m}) \le \check{\text{pen}}(m)$$

and we apply Lemma 6.1, which yields

$$\mathbb{E}[\|\check{\tau}_{\check{m}} - \tau\|^2] \le 3\|\tau - \tau_m\|^2 + 4\mathbb{E}[\check{\text{pen}}(m)] + 16c'\mathbb{E}^{1/2}\left[\frac{T^4 \mathbb{1}_{N_T \ge 1}}{N_T^6}\right].$$

This ends the proof of Theorem 3.1. □

6.8 Proof of Lemma 3.2

From (6), and under (A4) we have for $i \ge 1$, $R_{\widehat{T}_i^\Delta} = i$ *a.s.* and thus $\widehat{D}_{i+1}^\Delta = D_{i+1} + F_{\widehat{T}_i^\Delta - \Delta} - F_{\widehat{T}_{i+1}^\Delta - \Delta}$, where the three variables are independent by the renewal property. Under (A2) and for fixed time $t > 0$, the density of F_t does not depend on t and is given by τ_0 defined in (A2) (see *e.g.* formula (4.2.6) in Daley and Vere-Jones (2003)). Let $h : \mathbb{R}^+ \to \mathbb{R}$ be a bounded measurable function, we have

$$\mathbb{E}\big[h(F_{\widehat{T}_i^\Delta - \Delta})\big] = \sum_{j=1}^{\lfloor T\Delta^{-1} \rfloor} \mathbb{E}\big[h(F_{j\Delta - \Delta})\big|\widehat{T}_i^\Delta = j\Delta\big]\mathbb{P}(\widehat{T}_i^\Delta = j\Delta).$$

Moreover, for all $x \ge 0$ we have

$$\begin{aligned}\mathbb{P}\big(F_{j\Delta-\Delta} \le x\big|\widehat{T}_i^\Delta = j\Delta\big) &= \mathbb{P}\big(F_{j\Delta-\Delta} \le x\big|\exists i_0,\ T_{i_0} \in ((j-1)\Delta, j\Delta]\big)\\ &= \mathbb{P}\big(F_{j\Delta-\Delta} \le x\big|F_{j\Delta-\Delta} \le \Delta\big)\\ &= \frac{\int_0^{x\wedge\Delta}(1 - \int_0^y \tau(z)dz)dy}{\int_0^{\Delta}(1 - \int_0^y \tau(z)dz)dy} = \frac{x \wedge \Delta}{\Delta},\end{aligned}$$

where we used the dead-zone assumption (A4) to derive the last equality. Consequently, the variable $F_{j\Delta-\Delta}\big|\widehat{T}_i^\Delta = j\Delta$ has uniform distribution over $[0, \Delta]$. Then,

$$\mathbb{E}\big[h(F_{\widehat{T}_i^\Delta - \Delta})\big] = \sum_{j=1}^{\lfloor T\Delta^{-1} \rfloor} \int_0^\Delta \frac{1}{\Delta}h(y)dy\mathbb{P}(\widehat{T}_i^\Delta = j\Delta) = \int_0^\Delta \frac{1}{\Delta}h(y)dy,$$

which completes the proof. □

6.9 Proof of Proposition 3.2

To avoid cumbersomeness we work in the sequel as if the observations $(\widehat{D}_i^\Delta,\ 1 \le i \le R_T)$ were independent. Strictly, we should consider separately $(\widehat{D}_{2i}^\Delta,\ 2 \le 2i \le R_T)$ and $(\widehat{D}_{2i+1}^\Delta,\ 1 \le 2i+1 \le R_T)$, which are independent. But it is always possible in the sequel to split the sample, even if it means increasing slightly the constants.

First as $\widetilde{\tau}_m$ is in S_m, by Pythagoras Theorem we have

$$\begin{aligned}\|\widetilde{\tau}_m - \tau\|^2 = \|\tau - \tau_m\|^2 + \|\widetilde{\tau}_m - \tau_m\|^2 &= \|\tau - \tau_m\|^2 + \big\|\mathbf{G}_m(\Delta)^{-2}(\widetilde{\mathbf{b}}_m - \mathbf{b}_m)\big\|_{2,m}^2\\ &\le \|\tau - \tau_m\|^2 + \rho^2\big(\mathbf{G}_m(\Delta)^{-2}\big)\sum_{k=0}^{m-1}(\widetilde{b}_k - b_k)^2,\end{aligned}$$

where $\|.\|_{2,m}$ denotes the ℓ_2 euclidean norm of a vector of size m. Taking expectation and decomposing on the possible values of R_T, we are left to control

$$\mathbb{E}[(\widetilde{b}_k - b_k)^2] = \mathbb{E}\Big[\sum_{\ell=0}^{\infty} \mathbb{1}_{R_T=\ell}\Big(\frac{1}{\ell}\sum_{i=1}^{\ell}(\varphi_k(Y_i^\Delta) - \langle\varphi_k, f_\Delta(. - \Delta)\rangle)\Big)^2\Big].$$

We recover the same term as in the proof of Theorem 2.1, the same computations based on Lemma 6.2 lead to

$$\sum_{k=0}^{m-1}\mathbb{E}[(\widetilde{b}_k-b_k)^2]\leq 8m\mathbb{E}\Big[\frac{\mathbb{1}_{R_T\geq 1}}{R_T}\Big]+\mathbf{C}_1\|\tau\|^3\exp\Big(-\frac{\kappa'}{4\sqrt{2}\|\tau\|}\sqrt{m}\Big)+\mathbf{C}_2 m\sqrt{\mathbb{E}\Big[\frac{\mathbb{1}_{R_T\geq 1}}{R_T^4}\Big]}.$$

where $\mathbf{C}_1$ and $\mathbf{C}_2$ are given in (20). □

6.10 Proof of Theorem 3.2

Let $m_{\max}$ denote the maximum dimension m in $\widetilde{\mathcal{M}}_T$. Consider the vectors

$$\mathbf{t}=(a_0(t),\dots,a_{m_{\max}-1}(t))^T$$

in $\mathbb{R}^{m_{\max}}$, which are one-to-one related with functions t of $S_{m_{\max}}$ by $t=\sum_{j=0}^{m_{\max}-1}a_j(t)\varphi_j$. Vectors and functions spaces are denoted in the same way. If $\mathbf{t}$ is in S_m for $m\leq m_{\max}$ we have $a_m(t)=\ldots=a_{m_{\max}-1}(t)=0$. Let $[\mathbf{t}]_m$ be the m-dimensional vector with coordinates $(a_0(t),\dots,a_{m-1}(t))^T$. We also denote by $\langle\mathbf{u},\mathbf{v}\rangle_{\mathbb{R}^m}$ the vector scalar product in $\mathbb{R}^m$. Therefore, for $\mathbf{t}\in S_m$, thanks to the triangular form of $\mathbf{G}_m(\Delta)^{-2}$, we have

$$\langle\mathbf{t},\mathbf{G}_{m_{\max}}(\Delta)^{-2}\widetilde{\mathbf{b}}_{m_{\max}}\rangle_{\mathbb{R}^{m_{\max}}}=\langle[\mathbf{t}]_m,\mathbf{G}_m(\Delta)^{-2}\widetilde{\mathbf{b}}_m\rangle_{\mathbb{R}^m}.$$

Following the lines of the proof of Theorem 1 in Comte *et al.* (2016), and noticing that

$$\widetilde{\tau}_m=\arg\min_{\mathbf{t}_m\in S_m}\widetilde{\gamma}_T(t),\quad \widetilde{\gamma}_T(t)=\|\mathbf{t}_m\|^2_{\mathbb{R}^{m_{\max}}}-2\langle\mathbf{t}_m,\mathbf{G}_{m_{\max}}(\Delta)^{-2}\widetilde{\mathbf{b}}_{m_{\max}}\rangle_{\mathbb{R}^{m_{\max}}}$$

and

$$\widetilde{m}=\arg\min_{m\in\widetilde{\mathcal{M}}_T}\{\gamma_n(\widetilde{\tau}_m)+\widetilde{\mathrm{pen}}(m)\}$$

we obtain

$$\frac{1}{2}\|\widetilde{\tau}_{\widetilde{m}}-\tau\|^2\leq\frac{3}{2}\|\tau-\tau_m\|^2+\widetilde{\mathrm{pen}}(m)+4\sup_{\mathbf{t}\in S_{m\vee\widetilde{m}}}[\widetilde{\nu}_T(t)]^2-\widetilde{\mathrm{pen}}(\widetilde{m})$$

where

$$\widetilde{\nu}_T(\mathbf{t})=\langle\mathbf{t},\mathbf{G}_{m_{\max}}(\Delta)^{-2}(\widetilde{\mathbf{b}}_{m_{\max}}-\mathbf{b}_{m_{\max}})\rangle_{\mathbb{R}^{m_{\max}}}.$$

Now, define $\widetilde{p}_{R_T}(m,m')=\rho^2(\mathbf{G}_{m\vee m'}(\Delta)^{-2})p_{R_T}(m,m')$ with p_{R_T} defined in (28). Writing that

$$\mathbb{E}\left[\Big(\sup_{\mathbf{t}\in S_{m\vee\widetilde{m}}}[\widetilde{\nu}_{R_T}(t)]^2-\widetilde{p}_{R_T}(m,\widetilde{m})]\Big)_+\right]\leq\sum_{m'\in\widetilde{\mathcal{M}}_T}\mathbb{E}\left[\Big(\sup_{\mathbf{t}\in S_{m\vee m'}}[\widetilde{\nu}_{R_T}(t)]^2-\widetilde{p}_{R_T}(m,m')]\Big)_+\right]$$

and

$$\mathbb{E}\left[\Big(\sup_{\mathbf{t}\in S_{m\vee m'}}[\widetilde{\nu}_{R_T}(t)]^2-\widetilde{p}_{R_T}(m,m')]\Big)_+\right]\leq\rho^2(\mathbf{G}_{m\vee m'}(\Delta)^{-2})\mathbb{E}\left[\Big(\sup_{\mathbf{t}\in S_{m\vee m'}}[\nu_{R_T}(t)]^2-p_{R_T}(m,m')]\Big)_+\right]$$

we get the result. Indeed $\rho^2(\mathbf{G}_{m\vee m'}(\Delta)^{-2})\leq T$ in $\widetilde{\mathcal{M}}_T$ and the powers of R_T in the residual terms can be increased at the expense of slightly larger constants. □

6.11 Proof of Lemma 3.3

Recall that $g_k(\Delta)=\frac{1}{\Delta}\int_0^\Delta\varphi_j(x)dx$ and that $\Phi(x)=\int_x^{+\infty}\varphi_j(u)du$, we get $g_k(\Delta)=\frac{1}{\Delta}(\Phi_k(0)-\Phi_k(\Delta))$. Straightforward computations give

$$\int_0^{+\infty}\varphi_k(x)dx=\sqrt{2}\sum_{j=0}^{k}\binom{k}{j}\frac{(-1)^j}{j!}\int_0^{+\infty}(2x)^je^{-x}dx=\sqrt{2}\sum_{j=0}^{k}\binom{k}{j}(-2)^j=\sqrt{2}(-1)^k,$$

and (15) follows. For (16), we start from formula (38), yielding

$$\varphi_j(x) = \Phi_j(x) + 2\sum_{k=0}^{j-1} \Phi_k(x).$$

This formula implies (16) as

$$\varphi_{j+1} = \Phi_{j+1} + 2\sum_{k=0}^{j} \Phi_k = \Phi_{j+1} + \Phi_j + \underbrace{\Phi_j + 2\sum_{k=0}^{j-1} \Phi_k}_{=\varphi_j}. \quad \square$$

Appendix

A. Talagrand inequality

The result established below follows from the Talagrand concentration inequality given in Corollary 2 of Birgé and Massart [5].

Lemma 6.2. *Let $D_1, \dots, D_\ell$ be ℓ i.i.d. random variables and $\mathcal{F}$ a countable family of functions that are uniformly bounded by some constant b. Let $v = \sup_{t\in\mathcal{F}} \mathbb{E}[t(D_1)^2]$ and H_ℓ be such that $\mathbb{E}\big[\sup_{t\in\mathcal{F}} |\nu_\ell(t)|\big] \le H_\ell$. There exists a universal constant κ' such that, for any positive ε_ℓ, we have*

$$\mathbb{E}\Big[\Big(\sup_{t\in\mathcal{F}, \|t\|=1} \nu_\ell(t)^2 - 2(1+2\varepsilon_\ell)H_\ell^2\Big)_+^2\Big] \le 6\Big(\frac{2v}{\ell\kappa'}\Big)^2 \exp\Big(-\frac{\kappa'\ell\varepsilon_\ell H_\ell^2}{v}\Big) + 36\Big(\frac{2b}{\ell\kappa'}\Big)^4 \exp\Big(-\frac{\ell\kappa'\sqrt{\varepsilon_\ell H_\ell^2}}{\sqrt{2}b}\Big)$$

where $\nu_\ell(t) = \frac{1}{\ell}\sum_{j=1}^{\ell} \big(t(D_j) - \langle t, \tau\rangle\big)$, with the convention $\nu_0(t) = 0$, $\forall t \in \mathcal{F}$.

Proof of Lemma 6.2

The result is established using the Talagrand inequality and that for any positive random variable X we have $\mathbb{E}[X^2] = 2\int_0^\infty t\mathbb{P}(X \ge t)dt$. Denote by $X = \Big(\sup_{t\in S_m, \|t\|=1} \nu_\ell(t)^2 - 2(1+\varepsilon_\ell)H_\ell^2\Big)_+$, it follows that

$$\begin{aligned}
\mathbb{E}\big[X^2\big] &= 2\int_0^\infty t\mathbb{P}\Big(\sup_{t\in S_m, \|t\|=1} \nu_\ell(t)^2 \ge 2(1+2\varepsilon_\ell)H_\ell^2 + t\Big)dt \\
&= 2\int_0^\infty t\mathbb{P}\Big(\sup_{t\in S_m, \|t\|=1} \big|\nu_\ell(t)\big| \ge \sqrt{2(1+2\varepsilon_\ell)H_\ell^2 + t}\Big)dt \\
&\le 2\int_0^\infty t\mathbb{P}\Big(\sup_{t\in S_m, \|t\|=1} \big|\nu_\ell(t)\big| \ge \sqrt{(1+\varepsilon_\ell)}H_\ell + \sqrt{\varepsilon_\ell H_\ell^2 + \frac{t}{2}}\Big)dt.
\end{aligned}$$

We apply the Talagrand inequality (see *e.g.* Corollary 2 in Birgé and Massart [5]) with $\eta = (\sqrt{1+\varepsilon_\ell} - 1) \wedge 1$ and $\lambda_\ell = \sqrt{\varepsilon_\ell H_\ell^2 + t/2}$. We obtain, for κ' a universal constant,

$$\begin{aligned}
\mathbb{E}\big[X^2\big] &\le 6\int_0^\infty t\exp\Big(-\ell\kappa'\Big\{\frac{\varepsilon_\ell H_\ell^2 + t/2}{v} \wedge \frac{\sqrt{\varepsilon_\ell H_\ell^2 + t/2}}{b}\Big\}\Big)dt \\
&\le 6\int_0^\infty t\exp\Big(-\ell\kappa'\frac{\varepsilon_\ell H_\ell^2 + t/2}{v}\Big)dt + 6\int_0^\infty t\exp\Big(-\ell\kappa'\frac{\sqrt{\varepsilon_\ell H_\ell^2 + t/2}}{b}\Big)dt.
\end{aligned}$$

Next, we use that $\sqrt{\varepsilon_\ell H_\ell^2 + t/2} \geq \big(\sqrt{\varepsilon_\ell} H_\ell + \sqrt{t/2}\big)/\sqrt{2}$ to derive

$$\begin{aligned}
\mathbb{E}\big[X^2\big] \leq 6 \exp\Big(-\frac{\kappa' \ell \varepsilon_\ell H_\ell^2}{v}\Big) \int_0^\infty t \exp\Big(\frac{\kappa' \ell}{2v} t\Big) dt \\
+ 6 \exp\Big(-\frac{\ell \kappa'}{\sqrt{2}b}\sqrt{\varepsilon_\ell H_\ell^2}\Big) \int_0^\infty t \exp\Big(-\kappa' \frac{\ell \sqrt{t}}{2b}\Big) dt \\
= 6\Big(\frac{2v}{\ell\kappa'}\Big)^2 \exp\Big(-\frac{\kappa' \ell \varepsilon_\ell H_\ell^2}{v}\Big) + 36\Big(\frac{2b}{\kappa'\ell}\Big)^4 \exp\Big(-\frac{\ell\kappa'}{\sqrt{2}b}\sqrt{\varepsilon_\ell H_\ell^2}\Big).
\end{aligned}$$

Which is the desired result. □

B. Additional Numerical results

We present hereafter the numerical results corresponding to the distributions presented in Section 4. Tables 3-5 correspond to the comparison of the continuous time and the naive procedures and Tables 6-8 to the comparison of the continuous time and the dead-zone procedures. In all the tables, the lines $\mathbb{L}_2$ correspond to the value of mean square errors, $\overline{m}$ to the mean of the selected dimensions. All standard deviations are given in parenthesis.

T		$\Delta = 0$	$\Delta = 0.01$	$\Delta = 0.1$	$\Delta = 0.3$
500	$\overline{m}^3\Delta^2$		0.02	1.21	5.06
	$\mathbb{L}_2$	$3.64 \cdot 10^{-3}$ $(2.83 \cdot 10^{-3})$	$3.82 \cdot 10^{-3}$ $(3.05 \cdot 10^{-3})$	$10.87 \cdot 10^{-3}$ $(4.12 \cdot 10^{-3})$	$28.90 \cdot 10^{-3}$ $(2.69 \cdot 10^{-3})$
	$\overline{m}$	6.05 (1.12)	5.91 (1.22)	4.94 (2.35)	3.83 (0.52)
1000	$\overline{m}^3\Delta^2$		0.03	3.72	5.93
	$\mathbb{L}_2$	$2.06 \cdot 10^{-3}$ $(1.06 \cdot 10^{-3})$	$2.15 \cdot 10^{-3}$ $(1.12 \cdot 10^{-3})$	$8.41 \cdot 10^{-3}$ $(3.10 \cdot 10^{-3})$	$28.71 \cdot 10^{-3}$ $(1.65 \cdot 10^{-3})$
	$\overline{m}$	6.65 (0.89)	6.66 (1.13)	7.19 (2.65)	4.04 (0.30)
5000	$\overline{m}^3\Delta^2$		0.09	14.58	5.87
	$\mathbb{L}_2$	$0.75 \cdot 10^{-3}$ $(0.37 \cdot 10^{-3})$	$0.68 \cdot 10^{-3}$ $(0.37 \cdot 10^{-3})$	$7.23 \cdot 10^{-3}$ $(0.79 \cdot 10^{-3})$	$28.50 \cdot 10^{-3}$ $(0.78 \cdot 10^{-3})$
	$\overline{m}$	8.63 (1.96)	9.76 (1.94)	11.34 (0.63)	4.02 (0.17)

Table 3: Results for τ following a $|\mathcal{N}(1, \frac{1}{2})|$.

T		$\Delta = 0$	$\Delta = 0.01$	$\Delta = 0.1$	$\Delta = 0.3$
500	$\overline{m}^3\Delta^2$		0.06	6.19	42.89
	$\mathbb{L}_2$	$13.68 \cdot 10^{-3}$ $(3.96 \cdot 10^{-3})$	$13.82 \cdot 10^{-3}$ $(3.86 \cdot 10^{-3})$	$14.18 \cdot 10^{-3}$ $(3.66 \cdot 10^{-3})$	$16.27 \cdot 10^{-3}$ $(5.00 \cdot 10^{-3})$
	$\overline{m}$	8.75 (1.33)	8.70 (1.29)	8.52 (1.17)	7.81 (0.68)
1000	$\overline{m}^3\Delta^2$		0.12	10.25	46.67
	$\mathbb{L}_2$	$8.59 \cdot 10^{-3}$ $(3.47 \cdot 10^{-3})$	$8.61 \cdot 10^{-3}$ $(3.47 \cdot 10^{-3})$	$9.63 \cdot 10^{-3}$ $(4.00 \cdot 10^{-3})$	$14.50 \cdot 10^{-3}$ $(1.41 \cdot 10^{-3})$
	$\overline{m}$	10.62 (1.35)	10.60 (1.34)	10.08 (1.39)	8.03 (0.32)
5000	$\overline{m}^3\Delta^2$		0.40	19.21	48.18
	$\mathbb{L}_2$	$3.23 \cdot 10^{-3}$ $(0.78 \cdot 10^{-3})$	$3.27 \cdot 10^{-3}$ $(0.76 \cdot 10^{-3})$	$5.46 \cdot 10^{-3}$ $(1.62 \cdot 10^{-3})$	$13.81 \cdot 10^{-3}$ $(1.60 \cdot 10^{-3})$
	$\overline{m}$	15.96 (1.72)	15.86 (1.66)	12.43 (1.81)	8.12 (0.59)

Table 4: Results for τ following a $5 \times \mathcal{B}(6, 3)$.

T		$\Delta = 0$	$\Delta = 0.01$	$\Delta = 0.1$	$\Delta = 0.3$
500	$\overline{m}^3\Delta^2$		0.04	3.74	31.16
	$\mathbb{L}_2$	$7.43 \cdot 10^{-3}$	$7.47 \cdot 10^{-3}$	$7.35 \cdot 10^{-3}$	$10.42 \cdot 10^{-3}$
		$(2.67 \cdot 10^{-3})$	$(2.67 \cdot 10^{-3})$	$(2.60 \cdot 10^{-3})$	$(3.36 \cdot 10^{-3})$
	$\overline{m}$	7.43 (1.14)	7.41 1.20)	7.21 (0.69)	7.02 (0.19)
1000	$\overline{m}^3\Delta^2$		0.06	4.46	31.08
	$\mathbb{L}_2$	$5.30 \cdot 10^{-3}$	$5.31 \cdot 10^{-3}$	$5.82 \cdot 10^{-3}$	$9.12 \cdot 10^{-3}$
		$(2.02 \cdot 10^{-3})$	$(2.00 \cdot 10^{-3})$	$(1.53 \cdot 10^{-3})$	$(1.94 \cdot 10^{-3})$
	$\overline{m}$	8.49 (1.83)	8.47 (1.83)	7.64 (1.25)	7.02 (0.12)
5000	$\overline{m}^3\Delta^2$		0.18	11.76	30.88
	$\mathbb{L}_2$	$1.48 \cdot 10^{-3}$	$1.50 \cdot 10^{-3}$	$2.28 \cdot 10^{-3}$	$8.07 \cdot 10^{-3}$
		$(0.59 \cdot 10^{-3})$	$(0.57 \cdot 10^{-3})$	$(0.76 \cdot 10^{-3})$	$(0.74 \cdot 10^{-3})$
	$\overline{m}$	12.30 (2.11)	12.19 (2.06)	10.56 (0.68)	7.00 (0.03)

Table 5: Simulation results for τ following a $\left(0.4\mathcal{G}(2, \frac{1}{2}) + 0.6\mathcal{G}(16, \frac{1}{4})\right) \times \frac{8}{5}$ distribution.

T		$\Delta = 0$	$\Delta = 0.01$	$\Delta = 0.1$	$\Delta = 0.5$	$\Delta = 0.75$
500	$\mathbb{L}_2$	$5.89 \cdot 10^{-3}$	$7.30 \cdot 10^{-3}$	$7.43 \cdot 10^{-3}$	$10.13 \cdot 10^{-3}$	$14.63 \cdot 10^{-3}$
		$(3.48 \cdot 10^{-3})$	$(7.30 \cdot 10^{-3})$	$(7.43 \cdot 10^{-3})$	$(5.35 \cdot 10^{-3})$	$(10.57 \cdot 10^{-3})$
	$\overline{m}$	7.90 (0.68)	6.51 (1.03)	6.48 (1.05)	5.74 (1.30)	4.91 (1.15)
1000	$\mathbb{L}_2$	$4.96 \cdot 10^{-3}$	$5.02 \cdot 10^{-3}$	$5.02 \cdot 10^{-3}$	$5.69 \cdot 10^{-3}$	$7.49 \cdot 10^{-3}$
		$(2.28 \cdot 10^{-3})$	$(1.38 \cdot 10^{-3})$	$(1.34 \cdot 10^{-3})$	$(2.36 \cdot 10^{-3})$	$(3.11 \cdot 10^{-3})$
	$\overline{m}$	8.18 (0.39)	7.07 (0.30)	7.07 (0.28)	6.86 (0.49)	6.40 (0.66)
5000	$\mathbb{L}_2$	$4.99 \cdot 10^{-3}$	$4.12 \cdot 10^{-3}$	$4.13 \cdot 10^{-3}$	$4.50 \cdot 10^{-3}$	$4.94 \cdot 10^{-3}$
		$(0.94 \cdot 10^{-3})$	$(0.61 \cdot 10^{-3})$	$(0.60 \cdot 10^{-3})$	$(0.48 \cdot 10^{-3})$	$(0.35 \cdot 10^{-3})$
	$\overline{m}$	8.98 (0.16)	7.95 (0.21)	7.94 (0.23)	7.41 (0.49)	7.02 (0.15)

Table 6: Simulation results for τ following a $|\mathcal{N}(1, \frac{1}{2})|$ under the dead-zone assumption ($\eta = 1$).

T		$\Delta = 0$	$\Delta = 0.01$	$\Delta = 0.1$	$\Delta = 0.5$	$\Delta = 0.75$
500	$\mathbb{L}_2$	$9.30 \cdot 10^{-3}$	$16.25 \cdot 10^{-3}$	$16.30 \cdot 10^{-3}$	$22.49 \cdot 10^{-3}$	$45.01 \cdot 10^{-3}$
		$(4.41 \cdot 10^{-3})$	$(6.49 \cdot 10^{-3})$	$(6.54 \cdot 10^{-3})$	$(12.18 \cdot 10^{-3})$	$(14.14 \cdot 10^{-3})$
	$\overline{m}$	9.56 (1.24)	7.86 (0.74)	7.85 (0.75)	6.77 (1.10)	5.29 (0.45)
1000	$\mathbb{L}_2$	$5.74 \cdot 10^{-3}$	$11.60 \cdot 10^{-3}$	$11.66 \cdot 10^{-3}$	$15.36 \cdot 10^{-3}$	$23.90 \cdot 10^{-3}$
		$(3.27 \cdot 10^{-3})$	$(6.08 \cdot 10^{-3})$	$(6.09 \cdot 10^{-3})$	$(3.58 \cdot 10^{-3})$	$(6.90 \cdot 10^{-3})$
	$\overline{m}$	11.67 (1.37)	9.41 (1.52)	9.39 (1.52)	7.98 (0.22)	6.00 (0.33)
5000	$\mathbb{L}_2$	$1.91 \cdot 10^{-3}$	$2.19 \cdot 10^{-3}$	$2.21 \cdot 10^{-3}$	$6.43 \cdot 10^{-3}$	$14.93 \cdot 10^{-3}$
		$(0.67 \cdot 10^{-3})$	$(0.76 \cdot 10^{-3})$	$(0.78 \cdot 10^{-3})$	$(1.20 \cdot 10^{-3})$	$(1.46 \cdot 10^{-3})$
	$\overline{m}$	17.30 (1.87)	15.33 (1.13)	15.25 (1.05)	10.99 (0.13)	8 (0)

Table 7: Simulation results for τ following a $5 \times \mathcal{B}(6, 3)$ under the dead-zone assumption ($\eta = 1$).

T		$\Delta = 0$	$\Delta = 0.01$	$\Delta = 0.1$	$\Delta = 0.5$	$\Delta = 0.75$
500	$\mathbb{L}_2$	$14.76 \cdot 10^{-3}$ $(11.64 \cdot 10^{-3})$	$40.03 \cdot 10^{-3}$ $(29.77 \cdot 10^{-3})$	$40.91 \cdot 10^{-3}$ $(29.87 \cdot 10^{-3})$	$78.24 \cdot 10^{-3}$ $(18.51 \cdot 10^{-3})$	$84.35 \cdot 10^{-3}$ $(8.94 \cdot 10^{-3})$
	$\overline{m}$	13.01 (3.23)	8.36 (4.34)	8.22 (4.35)	2.90 (2.26)	2.11 (0.41)
1000	$\mathbb{L}_2$	$5.63 \cdot 10^{-3}$ $(3.34 \cdot 10^{-3})$	$10.19 \cdot 10^{-3}$ $(7.75 \cdot 10^{-3})$	$10.53 \cdot 10^{-3}$ $(7.98 \cdot 10^{-3})$	$31.61 \cdot 10^{-3}$ $(14.48 \cdot 10^{-3})$	$77.02 \cdot 10^{-3}$ $(14.03 \cdot 10^{-3})$
	$\overline{m}$	16.46 (1.22)	14.41 (2.15)	14.30 (2.21)	9.23 (2.14)	2.67 (1.02)
5000	$\mathbb{L}_2$	$2.85 \cdot 10^{-3}$ $(0.68 \cdot 10^{-3})$	$3.84 \cdot 10^{-3}$ $(1.11 \cdot 10^{-3})$	$3.88 \cdot 10^{-3}$ $(1.11 \cdot 10^{-3})$	$15.24 \cdot 10^{-3}$ $(2.91 \cdot 10^{-3})$	$29.53 \cdot 10^{-3}$ $(4.19 \cdot 10^{-3})$
	$\overline{m}$	17.82 (1.06)	16.49 (0.57)	16.45 (0.53)	11.02 (0.52)	9.88 (0.36)

Table 8: Simulation results for τ following a $\left(0.4\mathcal{G}(2, \frac{1}{2}) + 0.6\mathcal{G}(16, \frac{1}{4})\right) \times \frac{8}{5}$ under the dead-zone assumption ($\eta = 1$).

The mean of the number of observations is around 460 for $T = 500$, 930 for $T = 1000$, 4600 for $T = 5000$ in Table 3; around 147 for $T = 500$, 297 for $T = 1000$, 1497 for $T = 5000$ in Table 4; around 281 for $T = 565$, 297 for $T = 1000$, 2840 for $T = 5000$ in Table 5; 241 for $T = 500$, 485 for $T = 1000$, 2436 for $T = 5000$ in Table 6; around 112 for $T = 500$, 228 for $T = 1000$, 1151 for $T = 5000$ in Table 7; around 179 for $T = 500$, 361 for $T = 1000$, 1816 for $T = 5000$ in Table 8.

References

[1] A. Adekpedjou, E. A. Peña, and J. Quiton. Estimation and efficiency with recurrent event data under informative monitoring. *J. Statist. Plann. Inference*, 140(3):597–615, 2010.

[2] E. E. Alvarez. Estimation in stationary Markov renewal processes, with application to earthquake forecasting in Turkey. *Methodol. Comput. Appl. Probab.*, 7(1):119–130, 2005.

[3] J.-P. Baudry, C. Maugis and B. Michel. Slope heuristics: overview and implementation. *Stat. Comput.* **22**(2), 455–470, 2012.

[4] D. Belomestny, F. Comte and V. Genon-Catalot. Nonparametric Laguerre estimation in the multiplicative censoring model. Preprint MAP5 2016-1, 2016.

[5] L. Birgé and P. Massart. Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli*, 4(3):329–375, 1998.

[6] B. Bongioanni and J. L. Torrea. What is a Sobolev space for the Laguerre function systems? *Studia Math.*, 192(2):147–172, 2009.

[7] F. Comte, C., Duval, and V. Genon-Catalot. Nonparametric density estimation in compound Poisson processes using convolution power estimators. *Metrika* **77**, 163-183, 2014.

[8] F. Comte and V. Genon-Catalot. Adaptive Laguerre density estimation for mixed Poisson models. *Electronic Journal of Statistics* **9**, 1112-1148, 2015.

[9] F. Comte, C.-A. Cuenod, M. Pensky and Y. Rozenholc. Laplace deconvolution on the basis of time domain data and its application to dynamic contrast-enhanced imaging. To appear in *Journal of the Royal Statistical Society* **B**, DOI: 10.1111/rssb.12159, 2016.

[10] F. Comte and C. Dion. Nonparametric estimation in a multiplicative censoring model with symmetric noise. Preprint Hal and Preprint MAP5 2016-02, 2016. To appear in *Journal of Nonparametric Statistics.*

[11] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes. Vol. I.* Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2003. Elementary theory and methods.

[12] C. Duval. Density estimation for compound Poisson processes from discrete data. *Stochastic Process. Appl.*, 123(11):3963–3986, 2013a.

[13] C. Duval. Nonparametric estimation of a renewal reward process from discrete data. *Mathematical Methods of Statistics*, 22(1), 28-56, 2013b.

[14] I. Epifani, L. Ladelli, and A. Pievatolo. Bayesian estimation for a parametric Markov renewal model applied to seismic data. *Electron. J. Stat.*, 8(2):2264–2295, 2014.

[15] B. van Es, S. Gugushvili, and P. Spreij. A kernel type nonparametric density estimator for decompounding. *Bernoulli*, 13(3):672–694, 2007.

[16] J. Fan. On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.*, 19(3):1257–1272, 1991.

[17] R. D. Gill and N. Keiding. Product-limit estimators of the gap time distribution of a renewal process under different sampling patterns. *Lifetime Data Anal.*, 16(4):571–579, 2010.

[18] Gradshtein, I.S., Ryzhik, I.M. *Tables of integrals, series, and products.* Academic Press, New York, 1980.

[19] G. R. Grimmett and D. R. Stirzaker. *Probability and random processes*. Oxford University Press, New York, third edition, 2001.

[20] Y. Guédon and C. Cocozza-Thivent. Nonparametric estimation of renewal processes from count data. *Canad. J. Statist.*, 31(2):191–223, 2003.

[21] P. Hall and A. Meister. A ridge-parameter approach to deconvolution. *Ann. Statist.*, 35(4):1535–1558, 2007.

[22] M. Hoffmann and A. Olivier. Nonparametric estimation of the division rate of an age dependent branching process. *Stochastic Process. Appl.*, 126(5):1433–1471, 2016.

[23] T. Lindvall. *Lectures on the coupling method*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1992. A Wiley-Interscience Publication.

[24] G. Mabon. Adaptive deconvolution on the nonnegative real line, in revision, HAL preprint MAP5 2014-33, 2015.

[25] P. Massart. *Concentration inequalities and model selection*, volume 1896 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, With a foreword by Jean Picard.

[26] A. Meister. Deconvolution from Fourier-oscillating error densities under decay and smoothness restrictions. *Inverse Problems*, 24(1):015003, 14, 2008.

[27] G. K. Miller and U. N. Bhat. Estimation for renewal processes with unobservable gamma or Erlang interarrival times. *J. Statist. Plann. Inference*, 61(2):355–372, 1997.

[28] G. Soon and M. Woodroofe. Nonparametric estimation and consistency for renewal processes. *J. Statist. Plann. Inference*, 53(2):171–195, 1996.

[29] Y. Vardi. Nonparametric estimation in renewal processes. *Ann. Statist.*, 10(3):772–785, 1982.

[30] Y. Vardi. Multiplicative censoring, renewal processes, deconvolution and decreasing density: nonparametric estimation. *Biometrika*, 76(4):751–761, 1989.